

Computational Imaging and Automated Identification for Aqueous Environments

by

Nicholas C. Loomis

B.S., University of Nebraska - Lincoln (2005)

Submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

and the

WOODS HOLE OCEANOGRAPHIC INSTITUTION

June 2011

© Nicholas Loomis

All rights reserved.

The author hereby grants to MIT and WHOI permission to reproduce and
to distribute publicly paper and electronic copies of this thesis document
in whole or in part in any medium now known or hereafter created.

Signature of Author
Joint Program in Oceanography/Applied Ocean Science and Engineering
Massachusetts Institute of Technology
and Woods Hole Oceanographic Institution
29 April 2011

Certified by
George Barbastathis
Mechanical Engineering, MIT
Academic Advisor

Certified by
Cabell Davis
Biology Department, WHOI
Research Advisor

Certified by
Hanumant Singh
Applied Ocean Science and Engineering, WHOI
Thesis Supervisor

Accepted by
David E. Hardt
Graduate Officer, MIT Mechanical Engineering Dept.

Accepted by
James C. Preisig
Chair, Joint Committee for Applied Ocean Science and Engineering, WHOI

Computational Imaging and Automated Identification for Aqueous Environments

by

Nicholas C. Loomis

Submitted to the Joint Program in Oceanography/Applied Ocean Science and Engineering
Massachusetts Institute of Technology
and Woods Hole Oceanographic Institution
on 29 April 2011, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

Sampling the vast volumes of the ocean requires tools capable of observing from a distance while retaining detail necessary for biology and ecology, ideal for optical methods. Algorithms that work with existing SeaBED AUV imagery are developed, including habitat classification with bag-of-words models and multi-stage boosting for rockfish detection. Methods for extracting images of fish from videos of longline operations are demonstrated.

A prototype digital holographic imaging device is designed and tested for quantitative *in situ* microscale imaging. Theory to support the device is developed, including particle noise and the effects of motion. A Wigner-domain model provides optimal settings and optical limits for spherical and planar holographic references.

Algorithms to extract the information from real-world digital holograms are created. Focus metrics are discussed, including a novel focus detector using local Zernike moments. Two methods for estimating lateral positions of objects in holograms without reconstruction are presented by extending a summation kernel to spherical references and using a local frequency signature from a Riesz transform. A new metric for quickly estimating object depths without reconstruction is proposed and tested. An example application, quantifying oil droplet size distributions in an underwater plume, demonstrates the efficacy of the prototype and algorithms.

Academic Advisor: George Barbastathis
Title: Mechanical Engineering, MIT

Research Advisor: Cabell Davis
Title: Biology Department, WHOI

Thesis Supervisor: Hanumant Singh
Title: Applied Ocean Science and Engineering, WHOI

Acknowledgements

This thesis, and indeed my graduate education, has been a function of time, effort, and help from a multitude of people.

I owe a huge debt to my advisors and thesis committee for the effort they have put into my education. I thank Dr. Cabell Seal Davis for his continuous positivity, creativity, and interest in new methods. Cabell is responsible for my inclusion in the oceanographic community, my involvement in three adventurous cruises, and the majority of my marine microbiology knowledge. Dr. George Barbastathis graciously included me as part of his research group, giving me berth in his grad student offices and inviting me to participate as a member of his optics community. George has also provided perspective on academics, especially pertaining to optical systems. Dr. Hanumant Singh has been an energetic addition to my oceanic engineering education and has provided an especially helpful “big picture” approach. Dr. Jerry Milgram was my original thesis advisor before his retirement and introduced me to oceanic holography. It was his involvement that kickstarted the MIT-WHOI holography project and provided one of the central concepts:

“I saw what they were doing, and I said to myself, ‘Surely, we can do better than that!’” (Jerome Milgram, 2006)

My colleagues and coworkers in both the 3D Optical Systems Group and the Davis Lab have provided positive support and fruitful collaborations. Work with Dr. Weichang Li and Dr. Qiao Hu appears in Chapter 4; Dr. Se Baek Oh and Dr. Laura A. Waller shared an interest in phase-space methods, and their discussions led to developing the Wigner analysis in Chapter 3; experiments with Lei Tian laid the foundation for some of the ideas in Chapter 4; Dr. José Domínguez-Caballero was instrumental in helping develop the original benchtop systems that later became the prototype of Chapter 3. My gratitude extends to the entirety of both research labs.

Course work at MIT and WHOI was helpful throughout this thesis. In particular, course materials and projects from Dr. Antonio Torralba (computer vision, detection, and boosting), Dr. Frédo Durand (computational imaging, segmentation, and probability), Dr. Rosalind Picard (machine learning, pattern recognition, and probability), Dr. Alan

Edelman (parallel computing and algorithms), and Dr. Barbastathis (optical systems) are reflected directly, in several cases extended after the course to become complete sections of this thesis.

The MIT Museum has provided unique outreach opportunities for the ideas presented in this thesis. Seth Riskin originally involved our lab in holography activities at the MIT Museum, then Kurt Hasselbalch later proposed and guided the development of an interactive display on plankton holography. The computational approaches I learned for the museum display led to further scientific development and spurred the use of GPUs for processing; Section 3.3.5 is a direct consequence, as is the majority of the data processing performed throughout this thesis.

In reference to data processing, Matlab/MEX and NVIDIA's CUDA have been daily cornerstones for which I am continually thankful.

Family and friends have provided their own form of support and a life away from the computer screen. The MIT Cycling Team in particular was instrumental during my years, thanks especially to the cyclo-freakin-cross team, road team, and track team, and I consider many of the MIT riders to be integral parts of my Boston family. These friends have offered encouragement and reminders that everyone has their own challenges to meet [209],[351],[411], and that spagglng would not be allowed no matter how hard the race seemed to be at the time.

Dr. Tim Gay, Mr. Robert Scarborough, and the illustrious Mr. Jake Winemiller were hallmarks during my early scientific career. Tim exposed me to the field of AMO, the reason I decided to pursue optics during grad school. Jake and Robert both encouraged creative experimentation and quantifiable engineering. Jake, now that I have a PhD, you know what I plan to start doing.

Toscanini's has been, quite literally, my office away from the office and home away from home. The baristas keep me awake and working with hot drinks and friendly conversation, and the atmosphere of the shop is ideal for my mental process. I owe Tosci's tremendously for their fantastic working environment and my mental sanity. In particular, Sections 2.1, 2.2.3, 3.2, 4.1.3, and 4.1.5 all were started or developed while sitting at the big table.

Finally, mad props go to Dr. José "Pepe" Domínguez-Caballero, the most influential

person in my scientific development while at MIT. Pepe taught me digital holography and worked one-on-one with me throughout the beginning of my grad student career. He continued to be involved as a fellow holographer, a mentor, and a friend, discussing ideas, encouraging active experimentation, maintaining a curiosity about optics, and offering enough challenges that I was never left in need of projects. Pepe is a living example of the drive to achieve and innovate that I see as the MIT spirit.

Funding was provided by:

- NOAA Grant #5710002014, “Multiscale digital imaging of ocean species”
- NOAA NMFS Grant #NA17RJ1223, “Development of operational capability for automatic detection and classification of fish and their habitats from SeaBED AUV underwater images and video observer monitoring of commercial fishing vessels”
- NSF Grant #OCE-0925284, “Quantification of *Trichodesmium* spp. vertical and horizontal abundance patterns and nitrogen fixation in the western North Atlantic”
- NOAA Grant #NA10OAR4170086, “Holocam in gliders”

Contents

1	Introduction and Background	10
1.1	Sampling methods and devices	12
1.1.1	Large scale measurements	12
1.1.2	Traditional plankton measurements	13
1.1.3	Modern plankton measurements	14
1.1.4	Combined systems	17
1.1.5	Challenges for microscale optical devices	17
1.1.6	Challenges for macroscale optical devices	19
1.2	Contributions and highlights	20
2	Traditional Imaging Methods	22
2.1	Habitat classification from textures	23
2.1.1	Bag of words for texture classification	26
2.1.2	Texture descriptors	28
2.1.3	Multiple models per class	34
2.1.4	Classification accuracy	36
2.1.5	Habitat area coverage	39
2.2	Detection of rockfish	40
2.2.1	Boosted detection	44
2.2.2	Shape pre-processing and spurious detection removal	46
2.2.3	Shape-based classification	53
2.3	Longline detection and linking	61

2.3.1	Detection of anomalous areas	62
2.3.2	Linking detections	68
2.3.3	Improvement suggestions	73
3	Digital Holographic Instrument Development: Lab to Ocean	74
3.1	Introduction to digital holography	77
3.1.1	Modeling motion during exposure	86
3.1.2	Minimum number of bits	95
3.2	Wigner transform analysis for spherical reference waves	98
3.2.1	Wigner distribution and optical formulation	99
3.2.2	Insights to space-bandwidth limits	101
3.2.3	Space-bandwidth product and sampling volume	108
3.2.4	Depth of field	110
3.2.5	Subtle lessons	111
3.3	DH prototype design	112
3.3.1	System components	115
3.3.2	Benchtop systems with oceanic samples	119
3.3.3	Prototype design	122
3.3.4	Power filtering for Bayer detectors	127
3.3.5	Software for interactive reconstructions	133
3.4	Particle field effects	134
3.4.1	Image degradation	150
4	Extracting Information and Images from Holograms	156
4.1	Detection, position estimation, and focus determination	157
4.1.1	Traditional focusing methods	159
4.1.2	Combined focus metrics	169
4.1.3	Metrics using machine learning	174
4.1.4	Object detection from raw holograms	181
4.1.5	Fast depth estimation	192
4.2	Accelerated computations on graphics processing units	206

4.3	Application: Oil droplet size distributions	214
4.3.1	Droplet models	215
4.3.2	Detection using focus metrics	218
4.3.3	Fast circle parameter estimation	222
4.3.4	Droplet classification	228
5	Future Work and Conclusions	235
5.1	Contributions	235
5.2	Applications and extensions	238
5.2.1	Automatic identification of plankton	238
5.2.2	Particle size distributions	239
5.2.3	Biological tracking	240
5.2.4	Digital multiplexing holography	242
5.2.5	Holography on gliders, drifters, and AUVs	246
5.2.6	Riesz-based focus metric	249
5.3	Future outlook	250

Chapter 1

Introduction and Background

Gathering information about biological activity in marine environments has historically been challenging. An immense volume of water, high pressures, mobility, and the range of size scales have all had an impact on our ability to collect data. The oft-quoted metric is that more is known about the surface of the moon than the ocean [329] – because the measurement is that much easier despite the literally astronomical distances.

Tools to meet the challenge of biological and ecological sampling in the ocean have been growing in ability. The original sampling devices, nets and hooks, could return rudimentary information about certain species in a rough area. Their simplicity belied the amount of work required back in the laboratory. Later nets which could include improved spatial information by opening or closing on cue were developed in the late 1800’s and are still used today albeit with electronics to control the spatial sampling [421]. Modern devices which take advantage of electronic sensors and microprocessors are able to shift the burden of observation to the device itself. Instruments which capture data remotely through sound and light have enormously expanded the volume that can be sampled. Given that the ocean is estimated to have 1.3 billion cubic kilometers of water [58] and a seafloor larger than 350 million square kilometers, the ability to reach further is especially critical.

An understanding of the biology and ecology of the oceans goes well beyond an academic curiosity. The immediately obvious connection is that the ocean provides food and sustenance for humans and animals alike: in 2008, fish and seafood provided 15% of the animal protein for 3 billion people. Some 90 million metric tons of fish and aquatic plants are

captured each year, and another 50 million metric tons harvested through aquaculturing. Fishing and aquaculture provide jobs and financial support (including dependents) for 540 million people, about 8% of the world population. Fisheries exports were worth a record \$102 billion (US dollars) in 2008 [120],[119]. Less obvious is the regulating effect that the ocean has on global temperature and chemistry.

Changes in marine biodiversity “increasingly [impair] the ocean’s capacity to provide food, maintain water quality, and recover from perturbations” [428]. Humans have had a visible impact on worldwide fish populations, decimating certain populations and significantly altering the natural balance of many others [304],[34]. Unfortunately, the effect of fish populations is highly coupled, affecting animals lower in the food chain and modifying the predator-prey balance [430]. Overfishing in particular has been a long-standing problem in human history. Correcting overfishing, when possible, takes decades to centuries to achieve a stable balance [171].

The presence of humans also alters the chemistry of coastal areas through pollution and chemical run-off, affecting the marine balance in less direct ways than fishing [171]. Even far from the coasts, changes to the atmosphere are taken up by the ocean as it absorbs carbon dioxide and various anthropogenic chemicals. For example, the ocean is estimated to have absorbed around half of the carbon dioxide released from the combustion of fossil fuels. This has lead to a dramatic reduction in seawater pH and dissolved carbonate, affecting both the plant and animal life that depends on precise acid levels and chemical balances [112],[301].

A number of marine taxa are also sensitive to changes in temperature, leading to observable changes in the biodiversity [382]. Changes in both global and local temperature, both natural and human generated, have the ability to affect these species [143],[346].

Regulatory checks and balances can help preserve the environment and protect the future of the fisheries [34]. The process naturally requires good data about the current state of critical factors and an accurate understanding of how decisions will affect future populations [60],[93]. This is one of the critical areas where ocean sampling and observation enter the picture: obtaining the basic biological and environmental data to inform policy and science.

The interconnectedness of marine ecology means that information is required on multiple scales, and that observations about one species can provide input about another [279],[271]. Plankton, the smallest denizens of the aquatic food chain, are a prime example. Changes to the plankton population can ripple through the entire chain over time [21],[78],[376],[425]. The concentration and health of various plankton can also serve as sensitive indicators for global temperature and various environmental impacts [38],[429],[162],[111]. Tracking these changes generates predictions for the various domains that plankton populations impact – primarily animal, plant, and chemical [154],[22].

1.1 Sampling methods and devices

Gathering the fundamental data about plankton and fish populations, habitats, and the state of the oceanic environment again returns the discussion to the sampling problem. The ideal sampling instrument would be able to operate over a wide range of biological sizes, discriminate between species, collect data throughout a large volume of water, provide the 3D locations of all the objects in the volume, operate throughout the full ocean depth, and include information about the physical environment (salinity, temperature, currents, particulate matter, chemistry, etc.) – all without disturbing the subjects under study and thus ensuring an accurate measurement. These goals are by no means an exhaustive listing of “ideal”, but provide achievable goals with which to compare the various methods of collecting information in a marine environment. Several of the commonly used instruments for both large scale and planktonic sampling are worth reviewing with these goals in mind.

1.1.1 Large scale measurements

At the largest scale, satellites such as the SeaWiFS¹ can provide multispectral information about the upper layer (centimeters to a few meters) of the ocean based on backscattered light. Phytoplankton concentrations (including the ability to discriminate between a few dominant species), chlorophyll content, and size parameters can be correlated to colorimetric measurements [9],[68],[437]. The spatiotemporal coverage of a satellite depends on its orbital

¹The SeaWiFS recently stopped collecting data in December of 2010.

track, so that analyses are over relatively long time scales (days to years) and lack precise lateral position information (order of kilometers).

Sound has the ability to travel long distances through water, allowing for long-range sensing and extremely large sampling volumes using sonar. The distance and resolution are coupled to the wavelength so that sonar is reliably capable of imaging large fish and returning information about biomass [421],[172]. However, to reliably use sonar, models for scattering need to be created and tested. Detecting fish in the water column is easier than near the bottom where strong reflections occur, obscuring the return signal.

Optical solutions provide high resolution and easily interpretable results. Cameras and strobes are regularly mounted onto autonomous underwater vehicles (AUVs) and remotely operated vehicles (ROVs), used by divers, towed behind a vessel, or lowered on cables. They have the ability to image large volumes of water and visually discriminate between fish species and other centimeter- to meter- sized objects. Capturing information about benthic environments is done regularly. Optical methods are dependent on illumination and low scattering. Obtaining quantitative results from image sets can be time consuming and challenging.

1.1.2 Traditional plankton measurements

The earliest scientific device, a plankton or fish net, sweeps through a volume of water behind a vessel. Nets can sample incredibly large volumes of water. Detailed microscopic analysis of the captured animals provides high specificity, including information about life stage, gut contents, and reproductive status. Biochemical analyses, including DNA extraction, can also be performed. The three downsides are that spatial information is lost or rough at best, counting the species by hand in the laboratory is laborious and requires the talents of an expert, and the animals are forcibly removed from their environments. Nets which open and close at specific depths, for pre-set times, or which respond to signals from a control line (either physical or electronic) improve the spatial specificity slightly [421],[420].

The Continuous Plankton Recorder (CPR) is a variation of the net-capture-observe philosophy. Instead of collecting plankton using a fixed cod end, water is filtered past a silk mesh that is slowly transferred between two spools. The position of plankton on the mesh is

combined with knowledge of the CPR's path to piece together the location. The CPR is also innovative is that it is attached to ships of opportunity as they cross the Atlantic shipping lanes and has been providing data about plankton and micronekton along the transit lines since 1931, one of the longest running experiments in plankton sampling history [298],[412].

Efforts to automatically analyze plankton captured by CPR-like devices and nets has resulted in the ZooScan system. The captured plankton are laid onto a specialized flatbed scanner and imported into a computer where they are recognized using various machine learning approaches. The ZooScan reduces the effort of a human expert in analyzing each planktonic sample [147].

Nets have several problems that makes them unsuitable for certain types of biologicals. In particular, fragile objects such as gelatinous animals, trichodesmium colonies, or larvacean houses are destroyed or significantly underestimated [299],[87]. There may also be problems with avoidance, as some motile zooplankton can sense the shear from an approaching net and escape its path [42].

Pumps can increase the water sampled in an area, and are especially useful for studying small-scale relationships. The objects must be immotile, so that pumps are more often used for phytoplankton, microzooplankton, and particulates [289]. Pumps, like nets, have the downside of destroying fragile particulates.

1.1.3 Modern plankton measurements

Sonar, especially high frequency or multi-beam/multi-frequency setups, has seen continued interest for measuring plankton distributions. Euphausiids (krill) and jellies with air voids reflect sound with greater efficiency and can be measured to a degree. There has also been work to estimate plankton biomass using sonar. The three biggest problems for sonar are that most plankton is small and soft and thus does not efficiently reflect sound, models can only account for general categories of plankton, and the exact sonic properties of the water need to be known to account for changes in the observed signal. The consensus is that sonar may give rough estimates of biomass in certain populations, but is not a suitable tool for determining species or genus, especially for scales less than a millimeter [121],[172],[413],[421].

Non-imaging solutions have also been proposed and used for plankton measurements. The Optical Plankton Counter (OPC) and its later cousin, the Laser Optical Plankton Counter (LOPC) project a light sheet through the water and measure the statistical distribution of light intensities on a set of photodiodes. The OPC/LOPC provides spherical equivalent diameters of objects between about 1.5 mm and 35 mm, thus providing a size spectra only [158],[159],[59]. The Laser *In Situ* Scattering and Transmissometry (LISST) [3] has also been used to estimate phytoplankton size distributions. Laser light is diffracted from a small volume and imaged by ring photodetectors. Similar to the OPC/LOPC, the LISST provides size distributions only and is sensitive to the diffraction pattern [186],[185].

Optical plankton devices have proliferated as cameras and electronics have advanced. The Video Plankton Recorder (VPR), CritterCam[®]², Underwater Video Profiler (UVP) [146], and ZOOVIS (and ZOOVIS-SC, for “self-contained”) are all examples of camera-strobe pairs that use direct area imaging. The VPR images microscopic objects with a long working distance and is designed to be towed at high speeds (up to 10 knots for basin-scale measurements). It has a strobe opposite the camera at a slightly oblique angle (a ring in later versions) and essentially captures dark field images in either monochrome (original VPR) or color (VPR-II) [85],[84],[86],[88]. The Critter-Cam used Schlieren imaging for phase imaging of microscopic animals with a long working distance [363]. The ZOOVIS instruments use sheet illumination with a thickness on the order of the depth of field, and the camera is situated to image side scattering. The ZOOVIS is designed to be lowered downwards so that plankton encounter the light sheet before nearing any mechanical components, reducing avoidance [30],[28],[374].

Line scan camera systems have also been created for imaging plankton. The Shadowed Image Particle Profiling Evaluation Recorder (SIPPER) [311],[299] and *In Situ* Ichthyoplankton Imaging System (ISIIS) [79] both image objects by recording the shadow projected onto a line scan camera as the device is towed through the water. The SIPPER is intended for smaller plankton while the ISIIS is for larger mesoplankton and nekton. Both systems de-

²The CritterCam[®] was developed through a National Geographic Society grant and refers to a ruggedized video camera that can be used to image animals in their natural habitats. Research using the planktonic version has been extremely limited since the 1990’s. However, the CritterCam[®] (or Crittercam) has been attached to various animals since then, including whale sharks, seals, and various baleen whales – all which have close connections to plankton.

pend on the camera to be towed to generate images and the resulting sample volume (and image distortion) is a function of the tow speed.

Systems for particle imaging include the FlowCytobot [266] and the Submersible Flow-CAM (available commercially from Fluid Imaging Technologies). These pump water through an intake tube into an imaging chamber fitted with microscope optics. Pumps are intended for use with immotile, infrangible particles between a few microns and about half a millimeter.

The final class of optical measurement devices to discuss here is holographic devices. These occupy an interesting niche between imaging and non-imaging, as the hologram is the diffraction pattern but is later reconstructed as an image. Notable devices include the drifting unit from Katz et al., the eHoloCam, a device from Jericho and Kruezer, and the recently released commercial LISST-HOLO. The Katz unit was designed to drift with currents just below the surface, capturing holographic video of plankton interacting within their natural environment [281]. Jericho and Kruezer intentionally image microplankton [179],[137], and there are questions about avoidance that have not been addressed. The devices from Katz and Jericho/Kruezer both appear to be demonstration units and have seen limited use in biological studies. The eHoloCam has potential for biological studies and has been used a limited number of times. Its optical design includes a Q-switched laser, so that the device is best used on powered platforms [366],[367]. Current work with the eHoloCam seems to have stalled since about 2008. Sequoia Scientific, the manufacturers of the LISST, released a holographic version of a particle profiler in 2010. The engineering is rudimentary but allows basic holographic images to be recorded and reconstructed [321],[253]. A more complete review of holographic devices and their capabilities is included in Section 3.3.

The operating characteristics of the various imaging systems are primarily engineering and implementation choices. For example, the depth range can be extended for each instrument by using larger housings and syntactic foam. Similarly, power systems and data storage can be modified with enough time, effort, and grant money.

Several excellent papers further review the state of plankton imaging and optical imaging within the ocean, and provide an extended discussion of the exact needs that the devices are attempting to meet [82],[88],[172],[93],[334],[173],[421]. A review paper from Kocak et

al. that discusses new techniques and methods in imaging may be particularly interesting for optical scientists [193].

1.1.4 Combined systems

Each of the individual systems already discussed has its benefits and specific measurement regimes. Both temporary and permanent combinations have been tried with success for specific types of missions. For example, VPRs mounted to AUVs such as JASON, REMUS, or ABE are able to autonomously map out 2D areas or 3D volumes with fine detail [134], and attempts have been made with ROVs to track zooplankton [302],[317]. Common probes such as conductivity/temperature/density (CTD) sensors and fluorometers have been incorporated into later redesigns such as the VPRII [88] and the ZOOVIS-SC [374]. The BIOMAPPER-II is a particularly wide-reaching system that combines a VPR, CTD, fluorometer, transmissometer, radiometers, cameras, and sonar into a single towed platform [413].

1.1.5 Challenges for microscale optical devices

Direct collection of plankton by nets, CPR, divers, or other similar methods all have the same bottleneck: the need to identify the sample contents. Experts have to painstakingly re-sample and examine the contents. As expected with direct examination, the species resolution is extremely high. Automated or semi-automated systems such as the ZooScan can help reduce the need for an expert but still requires sample preparation and hands-on lab work and have taxonomic resolutions similar to the *in situ* imaging systems [147]. The rate at which samples can be processed and identified makes nets limited in their coverage and has led to the current sparsity of global data coverage.

Optical devices for plankton are faced with the trade-off between depth of field (DOF) and resolution³: the depth is proportional to the square of the resolution (see Chapter 3 and

³Optical resolution is defined as the minimum separation in the object plane at which two points can be discerned as distinct objects [36],[145],[155]. This is a property of both the optical and sampling system. Unfortunately, the “resolution” quoted by a surprising number of authors in the device literature is the pixel size of a detector or the diameter of the smallest isolated point object they can visually observe. Comparisons of resolution and depth of field should be taken with a grain of seasalt.

Device	Resolution	DOF	Min. object
VPR ⁴	10 – 50 μm (meas.)	0.7 – 5 cm (meas.)	300 μm
CritterCam [®]	15 μm (measured)	50 mm (measured)	(unknown)
SIPPER	> 50 μm (pixel size)	96 mm (as reported)	500 μm
ISIIS	> 68 μm (pixel size)	20 cm (as reported)	1-2 mm
ZOOVIZ	50 μm (measured)	1 cm (illumination)	1 mm
HoloPOD	6 – 12 μm (meas.)	> 250 μm per slice	150 μm

Table 1.1: Optical sampling capabilities of popular plankton imagers. The resolutions and DOF are quoted as listed in the literature. Measured values for the VPR are based on placing a test object at various locations and judging the useful limits from images captured by the camera. The DOF for the CritterCam was done similarly, using instead a crossed reticle target and visual judgements. The ZOOVIZ DOF was quoted as the thickness of the light sheet illumination. The resolution and DOF of the HoloPOD is based on both theory and measurement, and is discussed in greater detail in Chapter 3. The minimum object size is based on reports from the literature regarding the smallest object that could be reliably identified by the authors.

Figure 3-1 in particular). For example, capturing images with a 50 μm lateral resolution results in a DOF of only 2.5 mm. Extremely good resolution also requires a high numerical aperture and thus a minimum lens diameter that grows linearly with the working distance (or, more precisely, with the inverse optical path length, Equation 3.24). The resolution, DOF, and minimum object size for the more popular optical devices which can identify plankton species are listed in Table 1.1, with resolution and depth of field as quoted in the literature. These values are considered the working values, determined by experimenting with the actual devices using different targets. The trend is for resolutions greater than 50 μm and limited measured DOFs, so that these instruments are primarily useful for larger plankton. Fast frame rate cameras are used to achieve the necessary sampling volumes. The digital holographic imaging device reported on in this thesis, the HoloPOD, is included as the final entry in the table. It was designed with a goal of imaging a large range of plankton sizes, 150 μm to 30 mm, with an extended sampling volume per hologram and a volume per unit time comparable to the other optical devices. Chapter 3 reports on the theory, design, and testing of the HoloPOD device.

The optical samplers showcase two other issues that are important for plankton science. The first is that the sampling should be quantifiable. The VPR calibrates its sampling volume by measuring point scatterers at locations distributed through the imaging volume

and setting a threshold on a focus metric [318]. The CritterCam does not have a well-defined depth range, instead relying on object images to be too defocused or too poorly lit outside the intended sample. The SIPPER and ISIIS systems both require good estimates of the flow velocity to calculate the imaged volume at any instant in time. The ZOOVIZ assumes that the sheet illumination has a sharp spatial cut-off and that objects outside the illumination are not imaged. On the other hand, the HoloPOD has an exact image volume. The second major issue is that of avoidance. Several motile zooplankton and micronekton are able to sense the shear from an approaching device and will attempt to escape, skewing the totals downward. The ZOOVIS and VPR are specifically designed to reduce flow effects by using an extended working distance, with fluid modeling performed on the VPRII to limit shear in the image volume to levels lower than the detection threshold for most copepods [88]. The SIPPER, on the other hand, funnels its samples through the center of a large duct-like area, potentially leading to significant avoidance. The HoloPOD has a long working distance and a small hydrodynamic footprint, significantly reducing the shear and avoidance concerns.

Quantifying the images captured by optical devices is another significant challenge. Plankton imaging devices have a well-defined goal and design, so that the number of methods and software is as numerous as the devices themselves. Examples include AutoDeck and Visual Plankton (VPR/VPRII), Pisces (SIPPER), ZooScan (nets) [147], ZooImage and PhytoImage (FlowCAM), and the Plankton Analysis System and Plankton Interactive Classification Tool (PAS and PICT, general plankton recognition) [246].

1.1.6 Challenges for macroscale optical devices

Cameras mounted on AUVs, ROVs, and carried by divers offer a vastly different set of conditions. Variation in the background, orientation, and deformable objects means that experts are often required for parsing the imagery into useful data. Estimating habitat coverage, for example, often involves randomly sampling portions of the imagery and classifying the observed points. The totals are then estimated from a small portion of the dataset. Similarly, counting fish species involves an observer searching through images and tallying the numbers of the specific fish of interest. Needless to say, this can be incredibly time consuming

and slow, especially for missions that capture hundreds to thousands of pictures per dive. Addressing the need for automated methods in marine imagery is the goal of Chapter 2. Methods of determining habitat types and detecting rockfish from a downward looking AUV camera can not only aid a human observer but provide starting points for additional data such as estimating the sizes of fish or correlating habitat with abundance that would be especially time consuming using standard approaches. Detection and grouping of fish as they are caught on longlines will also be presented, an above-the-water marine application that has bearing for protecting the fisheries below.

1.2 Contributions and highlights

The main goal of this thesis is to develop and analyze tools for use in detecting and identifying biologically relevant objects in aquatic environments. The majority of the effort is focused on automated methods that are computationally practical for the large datasets generated in oceanography. Good sampling practice is also stressed, with attempts to estimate and measure the error of various algorithms or predict the performance of a new holographic device for plankton imaging.

The first foray is working with traditional images captured by digital cameras. A bag-of-words model is shown to be particularly good at correctly identifying habitats in AUV imagery. Small image patches provide an optimized filter, and recognition rates are improved by computing an independent components analysis on the filter basis. A multistage detector for rockfish is created from the same dataset, and includes discussion about why the detector and its features perform as they do. Chapter 2 concludes with detection and grouping of fish caught during longline operations and recorded by low bandwidth webcams.

An improved digital holographic imaging device for use with *in situ* plankton measurements is presented in Chapter 3. Theory predicting how it performs under motion and with limited bit counts informs the engineering decisions. An analysis of the spatial and bandwidth limits of spherical reference holography is done using Wigner transform methods, providing a complete and demonstrably useful model for general in-line holography. The specific engineering variables and choices for the digital holographic unit are discussed,

testing performed, and a prototype unit constructed. Real-time software for reconstructing the holographic images is presented. Theory and simulations describing the effects of particle fields, such as observed in marine holograms, is discussed.

Computational algorithms for extracting information from holograms captured by the prototype are presented in the fourth chapter. These are especially important for digital holography, limited in widespread use a lack of suitable algorithms for general imaging and descriptions of their performance. Various focus metrics are presented with an emphasis on fast computations for large reconstruction volumes. A novel focus metric that uses local Zernike moments as edge detectors is presented. For holograms which do not require full reconstructions, methods are suggested for quickly estimating the lateral position and depth of objects within holograms. Two approaches are presented for detecting objects laterally, one that extends a summation kernel to spherical reference holography and another that applies local frequency estimates to find areas consistent with a holographic signature. A new depth estimator is proposed, based on a normalized spectral response, and is demonstrated to have excellent depth resolution and noise insensitivity. The performance of focus metrics, lateral detectors, and the depth estimator with real-world oceanic holograms is presented. The methods are then applied to locating and sizing oil droplets in the Gulf of Mexico during a recent oil spill.

The final chapter discusses a number of extensions and ideas based on the work presented in this thesis.

Chapter 2

Traditional Imaging Methods

Photography is arguably one of the best ways to record information from a distance about our complex world. It has a long list of benefits: photography has an incredibly high space-bandwidth product compared to other measurement methods, camera and lens systems are well understood and developed, there is a high degree of flexibility in the imaging (for example, passive or active lighting, different spectral ranges, color filters, use of digital sensors, post-processing methods, and temporal information through video), the resulting picture is easy for a human to interpret, and many photographic setups can be inexpensive and simple – all of which contributes to the popularity of photographic methods for both the lay audience and scientific studies. The advent of digital cameras and improved computational methods have further boosted the abilities of photography to the point where it is a ubiquitous tool for both science and everyday life.

One of the challenges with modern imaging is that pictures are easy to capture, so that a scientific deployment can involve hundreds or thousands of pictures. The burden on an educated observer to quantify the data in those images can be immense – and incredibly time consuming. Computational methods which can reliably replace an observer, or filter out the important information for an observer, have a number of useful benefits: the ability to make complicated measurements (e.g., computing area coverage or fractal dimension), returning results faster than a human, and possible implementation on a vehicle for *in situ* decisions as a few examples.

This chapter focuses on automated methods for extracting specific information from

oceanic images using image processing and machine learning. Two particular data sources are used as examples: seafloor images captured by a downward looking camera on an automated underwater vehicle (AUV), and a low resolution video camera watching fish on a longline as it is pulled into a boat. These sources differ from many others (i.e., traffic and surveillance cameras, product quality control on conveyor belts, or photography in urban environments) in that the relevant information rarely follows a preferential orientation and there is not a straight-forward generative model which describes the varied shapes of the animals in the images. The methods developed in this chapter have application beyond oceanic use, as the purpose is to create texture recognition, object detection, and similarity grouping which has enough flexibility to work on the particularly challenging class of aquatic habitats and animals, all valid for cases with more constraints such as man-made textures and objects.

The term “traditional imaging” is used here to denote a detector and lens combination designed to image a plane of the image space onto the sensor – no steps are taken to modify the imaging system for the specific task aside from stopping down an aperture or selecting a different lens. The primary goal is to work with the images created from traditional imaging systems purely from the computational side after capture.

2.1 Habitat classification from textures

Biological information about the seafloor is immediately useful for oceanic biologists, chemists, and ecologists [60],[312],[394],[249]. Seafloor data has secondary use in the fisheries, as many crustaceans, mollusks, and certain profitable fish are benthic during larval stages of their life – if not their entire lives. Information about reef and coral ecology, along with the species inhabiting those areas, can be used as sensitive indicators for changing temperatures and chemical balances in different parts of the ocean [312],[142],[279],[429].

Habitat discrimination and species identification requires a high level of details and a broad field of view. Both tasks benefit from reliable color information. The SeaBED autonomous underwater vehicle (AUV) is engineered to provide imagery that meets these goals: high quality, a large field of view, careful color correction, and a fast enough imaging

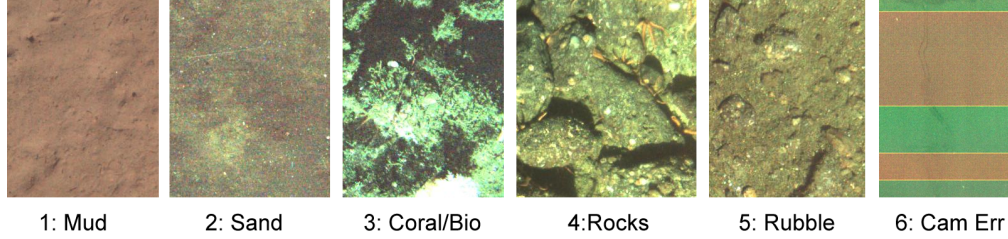


Figure 2-1: Samples of each of the six texture classes evident in the SeaBED images.

rate to acquire continuous swaths of data. During a single deployment, the SeaBED dives to depth, then cruises a few meters above the sea floor capturing several thousand images to an internal storage device [337],[338]. The image are retrieved when the AUV surfaces. Human experts can then analyze the images, searching out animals and interesting features (see, e.g., [384]).

Automated methods for extracting information from seafloor data has thus far been limited. The best examples are related to surface topology measurements: three-dimensional topology is made possible with multiple cameras and advanced algorithms for estimating pose and position [284],[283]. Describing the contents of the seafloor automatically is a different matter entirely. Unsupervised clustering may give computational results, but returns categories which have limited semantic meaning (see, e.g., [52],[285],[353], where multiple clusters correspond to the same category while other environmentally distinct categories are combined into the same cluster). Figure 2-1 shows samples acquired by the SeaBED and illustrates both the subtle visual differences possible (e.g., between *mud*, *sand*, and *rubble*) along with the gross differences (e.g., between *mud* and *coral/bio*¹) – both reasons why unsupervised learning is perhaps not the appropriate approach for habitat recognition.

The goal of this project is to use machine learning to perform habitat recognition through a texture recognition framework. The example data comes from ten SeaBED dives (Daisy Bank and Coquille Bank off the coast of Oregon and Santa Lucia Bank off the coast of Southern California; see [384] for location maps), a total of around 31,000 images (1.25 MPx color JPEGs, 10 GB total). The images were rigorously color corrected by the SeaBED

¹The class label coral/bio denotes rocky areas which have significant biological growth over the surface, or which have a coral-like appearance due to the biological activity. It does not necessarily indicate a true coral.

team, so that color can be used for classification of species and habitats both. Examples of the five predominant habitat classes, along with a sixth class to represent camera errors, are shown in Figure 2-1. The images have a significant range of size and orientations, though there is a bias for upward-facing shadows due to the fixed position of the strobe lights on the AUV. The habitats can also be mixed: sand coats the tops of large rocks or fills the area between rubble, for example. As mentioned, there is also a fine line between *rocks* and *rubble* and between *coral/bio* and *rocks*.

Texture classification has seen a number of new approaches in the past decade [440], including the use of “bag of words” models popularized by Varma and Zisserman [398],[399]. The bag of words (BoW) model compares the statistical distribution of filter responses for different textures, much like distinguishing between different documents by examining the frequency of characteristic word choices²[368]. It is particularly simple in that it disregards the spatial relationships between pixels, so that the filter response at one pixel is taken to be independent of its neighbor, and thus the distributions of the underlying random field need not be estimated. This in turn reduces the possible dictionary space and requires fewer training examples to estimate the distribution space. For natural textures without a preferred orientation (and thus a larger distribution space than oriented textures), this can be especially beneficial.

This section discusses recognition of seafloor textures using bag-of-words models, starting initially from the original Varma-Zisserman filterbank-based techniques and expanding out to incorporate multiple models per class label. An alternate view of filtering using image patches is explored, with links to optimal filter selection and transformation spaces. The resulting methods are tested for their classification accuracy, then used to measure areal habitat coverage across the full dataset – providing results to a problem which would be challenging and extremely time consuming for a human observer, but computationally tractable for a single desktop computer.

²The bag-of-words model uses many typographic terms based on its lexicographic foundation, the most notable here being a “dictionary”, or codebook, of the most common texture “words”.

2.1.1 Bag of words for texture classification

The bag of words model uses the statistics of an unordered collection of related elements to perform recognition [440],[368]. In the case of texture recognition, those related elements are the texture descriptors computed at each pixel in a digital image, termed “textons” by the computer vision community. A new texture is recognized by first determining the best texton to represent each pixel, then comparing the frequency of each texton against its expected frequency for known textures. The term “best” is intentionally vague, as its meaning will change depending on how the similarity between feature vectors is measured³, but is in general a measure of minimum distance.

Training is performed in two steps (Figure 2-2). In the first step, a series of filter responses are computed for each pixel in a set of training images for a single class, forming a feature vector at each pixel. The feature vectors are aggregated and quantized into representative clusters using k -means, with each cluster center representing a texton for that training class. The textons for all training classes are gathered into a *dictionary* of representative textons.

The second step of training uses the dictionary to estimate texton distribution models for each class. The feature vectors are again computed for each pixel in a training image, then each pixel is labeled with the dictionary texton which has the smallest distance to the feature vector. (If the filters are normalized to the same value, the response for each feature vector component is on the same order and a Euclidean distance can be used. A weighted Euclidean distance or a Mahalanobis distance, Equation 2.8, may be a better choice if the filters have different magnitudes [80],[236].) The frequency distribution of texton labels is then computed and becomes the model for that particular training image. A class model can be estimated by averaging together the models for each training image in that class if the models are similar enough, by using k -means or another clustering algorithm [106] to select a limited number of models if there is dissimilarity between models of the same training class, or by maintaining the entire collection of image models. The first two options have

³One simple example is a feature vector which includes components with different scales, such as local mean and local entropy. In that case, a Mahalanobis distance [236] may be more appropriate than a Euclidean distance.

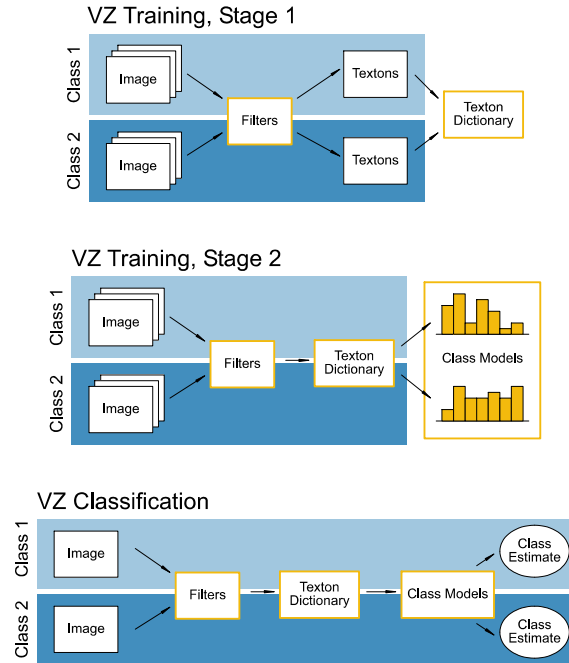


Figure 2-2: Steps in VZ texture recognition. In the first stage of training, sample images are passed through a set of filters and their responses clustered to create a set of representative textons for that class. The second stage computes class models based on the frequency of the different textons appearing in the training images. Classification is done by passing new images through the same filterbank and computing its distribution of texton responses; the class with the most similar texton distribution is selected as the sample's label.

the ability to remove or reduce the effect of outliers, while the last option can be sensitive to outliers (which may be desired in some cases) and requires all training images to be labeled correctly.

Recognition of textures is done using a similar method as the second training step. The texton labels are again computed for each pixel of the image of an unknown texture, then the texton distribution is computed. The distribution is compared against the class models and the class with the smallest χ^2 value is selected as the best estimate. The χ^2 distance is calculated as

$$\chi^2 = \sum_i \frac{(x_i - y_i)^2}{x_i + y_i} \quad (2.1)$$

for two discrete distributions x and y , where x_i is the value of the i^{th} bin of x [368].

A number of variations for BoW are immediately obvious: the k -means clustering during the dictionary creation step can be replaced by affinity propagation [128], k -nearest neighbors [106], or a hierarchical mean shift [278] with the ability to adjust the importance/similarity of individual textons; the χ^2 distance can be replaced by other distribution-distance measures such as a symmetric Kullback-Leibler [180], Bhattacharyya [4],[33] (which itself is directly related to the Matusita distance [4],[247]), or Kolmogorov-Smirnov metrics; the texton dictionary can be pruned to remove textons appearing in multiple classes; and so on. The interest here is in the overall method, and minor tweaking is left to future users. The remainder of this section will concentrate on using k -means for computationally efficient clustering, affinity propagation when selecting multiple models per class, and χ^2 for comparing models.

2.1.2 Texture descriptors

The traditional Varma-Zisserman (VZ) approach uses filter responses to represent a texture description. Their preferred filter bank is the MR8 bank, which includes eight filters: three sizes of bar filters, three sizes of edge filters, a Gaussian, and a Laplacian of a Gaussian. Each bar and edge filter is computed for multiple angles and the maximum response across the angles is used as that filter’s overall response [398]. Other filter banks are certainly possible; see [398] for descriptions of several types compared in their work. In the exploratory phase

for this work, a filter bank of scaled and oriented Gabor filters [198], one using Hu invariants [164], and one composed of local statistics (local mean, local variance, and scaled local entropy) were tried. The filter bank of local statistics gave reasonable results despite its ad-hoc nature and provides a baseline for comparison. The other filter banks gave poor results and were not explored further.

Recently, the idea of using image patches extracted directly from the texture images was proposed as a new feature vector. A small block of pixels around the pixel of interest is reshaped into a vector, normalized appropriately, and used directly to create the texton dictionary in the same way as a vector of filter responses [399]. The texton label assigned to each pixel is then the dictionary texton with the minimum Euclidean distance to that pixel’s image patch.

The patches used here were created by combining grayscale and color information. The image was first converted to grayscale, mean subtracted, and normalized to the standard deviation to remove intensity artifacts. Each $n \times n$ normalized intensity patch was reshaped into an $n^2 \times 1$ vector. Color information was included by appending a 3×1 vector of the mean values of the RGB color channels over the patch, made possible by careful color correction performed during the data acquisition. The RGB values range from $[0, 1]$, so that they have similar magnitudes as the normalized intensity information. The use of non-linear color spaces, color invariants [395],[396],[45], or a $(3n^2) \times 1$ vector which retains all of the data from each of the color channels are left for future study. Notably, non-linear color spaces such as HSV or HSL [383] would require a distance metric which incorporates the angular hue component.

The patch approach has several benefits. First, it does not require a specific filter bank, removing one level of obfuscation and experimentation. The patch textons may actually be better than arbitrarily selected filter banks as they are the result of using vector quantization with each texture, forming a compact set of exact representations [140]. Second, patches can act like a kernel method by increasing the dimensionality of the problem, possibly leading to better discrimination [106]. Third, as Varma and Zisserman point out [399], large-scale gradients or textures can be categorized by examining the histograms of the local gradients, so that much of the same information as in filter banks is present in patches.

Evidence that texton patches contain similar information to filters can be seen by examining the selected textons. Figure 2-3 shows an example texton dictionary selected for a set of 5×5 patches. A number of patches depict bars, edges, and corners with various orientations, similar to the MR8 filter bank, but with additional specificity for the scales and spatial frequencies present in the observed data. For comparison, the texton dictionary selected by Varma and Zisserman is shown in Figure 2-4, which includes a significant number of man-made textures. The VZ dictionary again contains a large number of bars, edges, and corners, though with a number of high-frequency stripes to accommodate the synthetically manufactured textures. Both of these dictionaries suggest that bars, edges, and corners are good representations of the information content in generic textures. Work from Torralba et al. suggests that this extends to generic images as well: they use a boosting algorithm to select patches (which are used in their work as filters) which provide good recognition and discrimination between a large number of object categories [385]. Their best filter patches are shown in Figure 2-5 – and include a number of bar, edge, and corners along with a few more specific filters for classes which are otherwise difficult to discriminate. The overall message is that patches can contain the same information as filter banks, while offering high specificity and the ability to generalize.

Dictionary textons are selected in BoW for each class alone, then aggregated together. This has the potential of generating redundant textons. Some dictionaries may also be linearly dependent, since patches span \mathbb{R}^{n^2+3} at most and dictionaries which contain more than $n^2 + 3$ elements are easy to generate. Two transforms to increase the discriminability and independence of the textons were considered: an eigenmode decomposition and an independent components analysis.

The eigenmode decomposition was computed by taking the singular value decomposition (SVD) of a set of dictionary textons. The singular vectors corresponding to non-zero singular values (a total of $n^2 + 3$ at most), termed “eigenpatches” when the SVD is applied to patch textons, are retained as an appropriate basis set for transforming patches into the shared eigenspace. The first training stage is modified by projecting the previously-determined dictionary textons into the eigenpatch basis to form a new, transformed dictionary. The second stage is performed by again extracting patches from images, then decomposing the

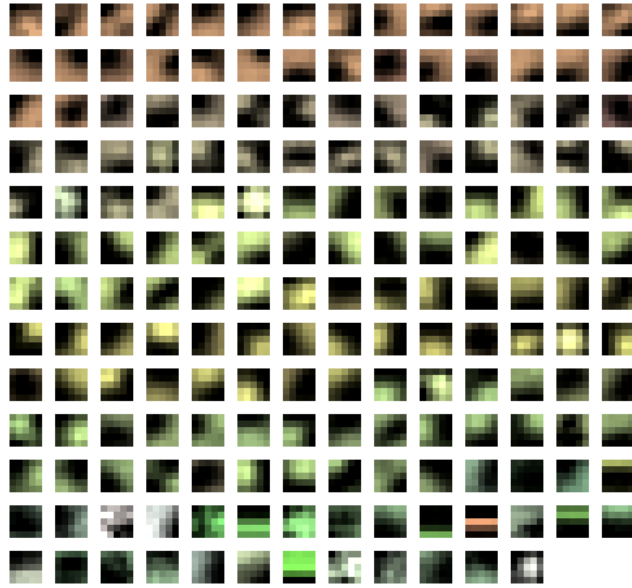


Figure 2-3: Patch textons selected by k -means for the habitat classification problem. The patches are 5×5 pixels each and there are 30 textons per class for a total of 180 dictionary textons. The textons flow from top to bottom in order of their ordinal class number (Figure 2-1).

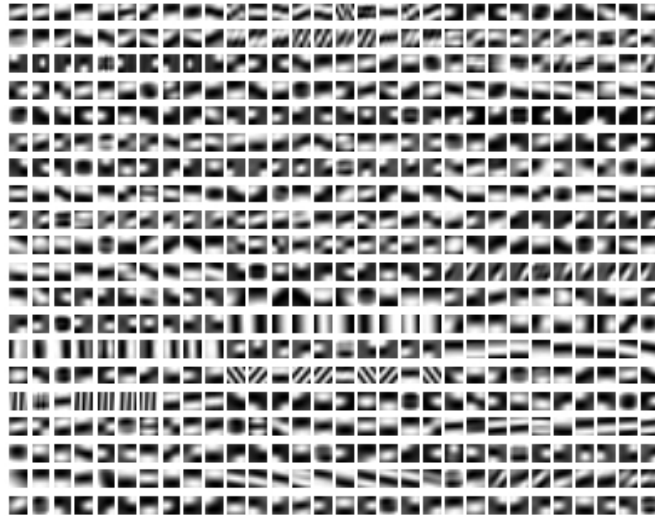


Figure 2-4: Dictionary patch textons selected for a collection of man-made textures; figure is from [399].

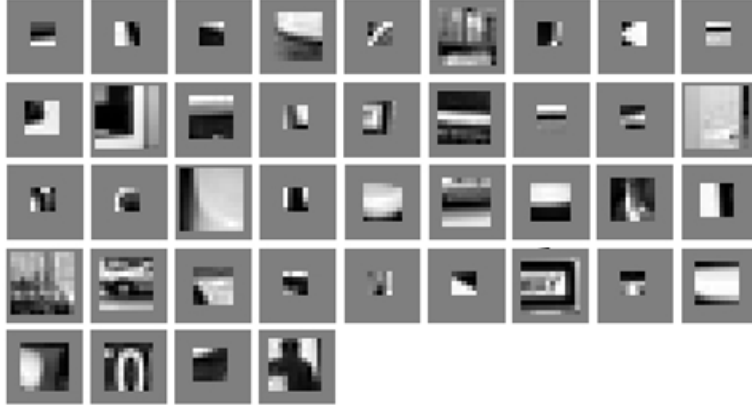


Figure 2-5: Patch filters selected by boosting for recognizing a large class of man-made objects and textures; figure edited from [385]. Note that the majority of the edges and bars have vertical, horizontal, or 45 degree orientations due to their origin from man-made objects – which is different from unoriented natural objects.

patches into the eigenpatch basis to form transformed feature vectors. These eigenpatch vectors are used with the transformed dictionary to create the class models. The recognition step similarly includes an eigenpatch transformation when computing models for the unknown texture.

An example set of eigenpatches corresponding to the patches in Figure 2-3 is depicted in Figure 2-6. (Colors may be inverted since the singular vectors have a sign ambiguity.) The first few eigenpatches depict bars and edges, similar to Figures 2-3, 2-4, and 2-5 and the MR8 filter bank – these are the basic building blocks which form the basis of many images. Higher spatial frequencies are reserved exclusively for the eigenpatches corresponding to the eigenvalues with smaller magnitude (higher indices). However, a significant amount of energy is spread into the higher-index eigenpatches (20% of the energy is in the last 14 of 28 patches), indicating that there may be useful discriminability in the higher eigenpatches. The problem is that these higher-index eigenpatches individually have small energy compared to the common low-index eigenpatches, making the difference between the transformed patches difficult to detect.

The second transformation attempts to find a more discriminable basis set by using an independent components analysis (ICA). The ICA finds a basis in which the data are less Gaussian and are thus closer to being statistically independent [170]. The resulting “ICA

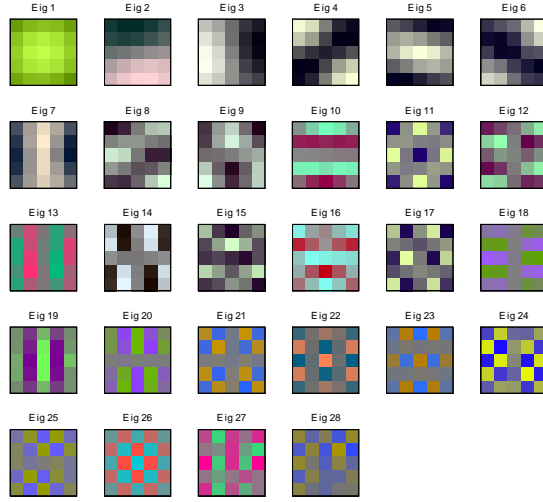


Figure 2-6: Eigenpatches for a set of 5×5 patches. A total of 28 patches are available here due to the additional three color components. Eigenpatches have been normalized for display, and colors may be inverted. The patches are formed by reshaping the intensity portion of the singular vector and applying multiplying by the mean “color”. The eigenpatches are shown in decreasing order of how much signal energy they represent.

patches” are, in one sense, more unique and thus give better discriminability. Figure 2-7 shows the ICA patches generated from the forward ICA transform⁴ corresponding to the patch dictionary of Figure 2-3 [169],[138]. (Color may be inverted; similar to eigenpatches, the ICA transform vectors do not include sign information.) The ICA patches tend to highlight small peaks or dips and, perhaps more importantly, where those peaks and dips appear in the patch: the shifts, such as between ICA patches 2, 4, and 8, differ in Fourier space by their phase ramps. (Similarly, consider ICA patches 10, 14, and 22). This suggests that the ICA patches are types of phase-space filters. The ICA patches are used like the eigenpatches, transforming image patches during the model generation stage of training and recognition.

⁴An ICA includes both a forward and inverse transform. The forward transform describes the underlying components which are used to generate the observed features, while the inverse maps observations back to the independent feature space. This is comparable to the U and V matrices of the SVD, where a matrix A is decomposed into $A = USV^T$.

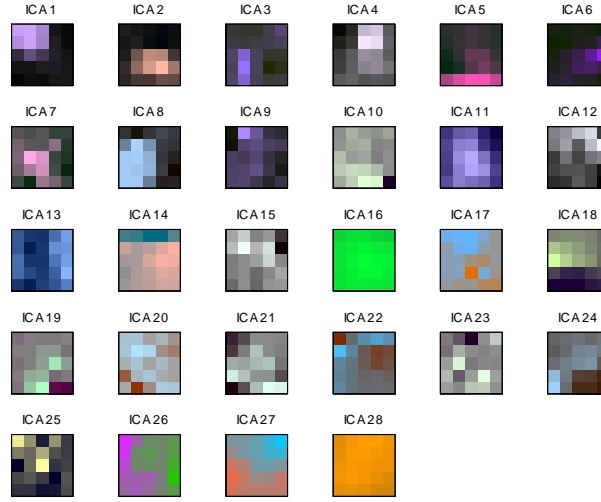


Figure 2-7: ICA patches computed for 5×5 pixel image patches; shown here are the forward ICA transforms.

2.1.3 Multiple models per class

Natural textures are particularly prone to have greater variation within each semantic class. For example, the difference between small boulders and large rubble is visually apparent, leading to distinct models for each component – but both have the same connotation for a biologist since they support the same set of species. There are also a number of images where the difference between rubble and small boulders is minor (or some rubble exists with a set of small boulders and vice versa), so that consistently labeling the images cleanly into two separate classes is difficult at best. The best solution for this case would be to include multiple models, at least one for small boulders and one for rubble, under the same *rocky* label.

There are two ways of creating multiple models per class. One is to retain a model for every image in the training set. This has the ability to map out a large feature space, assuming each of the training images has the correct label. Unfortunately, this approach can be sensitive to outliers, especially if models from one class overlap into the area of another class. Another issue is that many more training samples are required to adequately map out the feature space belonging to each class. Clustering models together can help alleviate

some of these issues: it reduces the effects of outliers and can select reasonable models with fewer samples. A caution with clustering is that it reduces the specificity of the feature space-to-class mapping. In this work, clustering was used to choose a limited number of relevant models for each class so that a smaller number of training samples could be used.

Clustering, or unsupervised learning, has a huge proliferation of methods. Already, k -means was discussed as a simple way of selecting specific numbers of clusters – if the number of clusters is known *a priori*. Besides needing to know the number of clusters, it has a potentially serious drawback: the cluster centers are taken as the mean of the cluster elements, so that cluster centers may not actually be members of the set (especially if the wrong number of clusters are used).

Affinity propagation (AP) is a new method which uses the similarity between elements to select a few elements which best exemplify the cluster characteristics. It has been shown to select better clusters than k -means in several cases, can better accommodate clusters with varied sizes, and can cluster based on non-standard similarity metrics [128],[252]. One of the reasons that AP is used in this work is that the clustering algorithm returns a measure of the net similarity of the clusters, $NS(p)$, and the number of clusters, $C(p)$, as a function of the initial clustering preferability, p . An AIC-like criterion [46],[7] is computed as

$$AIC' = 2C(p) - 2NS(p),$$

and the model clusters corresponding to the minimum AIC' are selected as the appropriate models for the training class. The original derivation from Akaike includes a logarithm of the likelihood function [7], which is replaced by NS here as a way to approximately measure the agreement of the data with the clustering. The scale value of two was selected experimentally to give reasonable clustering results. The similarity between clustering elements was computed using the χ^2 distance. Most training classes in the habitat data set resulted in one to four models per class.

2.1.4 Classification accuracy

The real-world seafloor habitats of Figure 2-1 from the SeaBED AUV dataset were tested for classification accuracy using statistical filters⁵, direct image patches, eigenpatches, and ICA patches. A total of 631 images with uniform class membership, as judged by a human expert, were selected randomly from the over 30,000 images in the example dataset and labeled with one of *mud*, *sand*, *coral/bio*, *rock*, *rubble*, or *camera-error* as the true class label. Representative textons were found by randomly selecting 10,000 textons from each class and using k -means to compute k cluster centers; the k clusters were aggregated from each class to form a complete dictionary with a total of $6k$ texton elements. Patch transforms were applied to the entire dictionary. Methods which used a single model per class formed the model from the 10,000 textons used to initially form the dictionary since they represented a random selection drawn throughout the class. The multiple models per class case used twenty images randomly selected from each class to form an initial set of models. The texton frequencies were computed for each of the twenty images and affinity propagation used to select appropriate class models. Images corresponding to the models selected to represent the class were removed from the test set. The confusion matrix and true positive rate were recorded for each experiment. Throughout this sub-section, k is the number of textons per class used when creating the dictionary, n is the number of pixels per edge in an image patch (i.e., the patch is sized $n \times n$), a “-S” following a method name denotes that the results were computed using a single model per class (e.g., “Eigenpatches-S”) and a “-M” denotes the use of multiple models per class.

Overall classification rates for the ad-hoc collection of statistical filters (local mean, local standard deviation, and scaled local entropy) is shown in Table 2.1. The statistical filters gave better performance than either the MR8 or Gabor filter banks despite its contrived nature. The table is shown for the sake of providing a baseline: for the SeaBED images, overall recognition rates of 85-89% are possible with the right set of filters. (Gabor and MR8 were in the 65% to 80% range.) The goal for patch methods is then to improve the

⁵Additional testing with MR8 and Gabor filter banks was done, but is not reported here: the results were poor and not particularly illuminating. The statistical filters themselves are reported here for the sake of providing a baseline.

k	Filter-S	Filter-M
10	85.9%	86.7%
30	89.1%	84.6%

Table 2.1: True positive results obtained using statistical filters.

n	k	Patch-S	Patch-M	Eigenpatch-S	ICA patch-S	ICA patch-M
3	10	83.0%	81.1	84.3	91.3	90.6
3	30	84.6	87.9	84.8	92.1	93.0
5	10	84.6	75.1	84.9	92.2	92.9
5	30	86.7	78.9	86.5	93.5	92.1
7	10	82.3	78.7	83.0	92.2	92.6
7	30	86.4	78.8	86.2	93.5	92.9
9	10	82.4	74.6		92.2	92.7
9	30	85.6	80.7		92.9	92.9
11	10	83.5	69.0			
11	30	84.0	77.6			

Table 2.2: Recognition results for various patch-based methods. Values are the overall true positive rates.

recognition rates significantly over the rates from filter methods.

A selection of results for patches are shown in Table 2.2 (additional experiments with larger n and additional patch-based methods, such as MRFs, such as in [399], is not shown here). The direct use of patches, Patch-S, has comparable results to the statistical filters and better results than the MR8 filter bank, supporting the idea that patches have enough representational power to rival filters. This in itself is useful as it reduces the work required by an expert in finding and selecting a specific filter bank. Patches with multiple models have significantly worse performance due to confusion between the *coral/bio*, *rock*, and *rubble* classes. These classes would have had at least one model close to a model from the other classes, a drawback of having too many models to span a small feature space. The use of eigenpatches did not significantly increase the discriminability above that of direct patches. ICA patches, both single and multiple model, increased the recognition rates markedly, by 6-10% over direct patches. The statistical independence generated by the ICA transform appears to boost the discriminability and is worth the additional computational effort in computing the ICA (a slow process for high-dimensional data but done only once during training) and applying the transform to each patch.

Statistic	Patch-S	Patch-M	Eigenpatch-S	ICA patch-S	ICA patch-M
Mean TP Rate	84.4%	78.2	85.0	92.7	92.7
Std. Dev.	1.48%	4.92	1.28	0.69	0.76
Best (n, k)	86.7% (5,30)	87.9 (3,30)	86.5 (5,30)	93.7 (5,60) (7,60)	94.8 (9,60)
Total Number	16	10	6	24	24

Table 2.3: Statistical summary of experiments with patch-based methods. Mean, standard deviation, and best recognition rate are percentages.

	Mud	Sand	Coral/Bio	Rock	Rubble	Error	TP Rate
Mud	97						100%
Sand	3	272			3		97.8%
Coral/Bio			26	2	2		86.7%
Rock	2		3	101	11		86.3%
Rubble		1		2	85		96.6%
Error	2	1			0	18	85.7%

Table 2.4: Confusion matrix for the best classifier found during experimentation. The true class is listed in the rows, the estimated class in the columns. Zeros are left blank for clarity of comparison.

The mean classification rate across all (n, k) combinations is shown in Table 2.3 for each method. The statistics assume that the major contribution is from the method as opposed to the patch size or dictionary size and provide rudimentary evidence that the ICA patches are indeed better than the other patch methods. For the direct patch, sizes of up to $n = 11$ were used for a total of sixteen experiments. For the ICA patches, a total of 24 experiments were tried with dictionaries ranging from ten to sixty textons per class in steps of ten.

The confusion matrix corresponding to the best recognition rate over all the experiments, 94.8% overall true positive, is shown in Table 2.4. The confusion matrix corresponds to ICA patches-M with $n = 9$ and $k = 60$. The greatest source of confusion is between *coral/bio*, *rocks*, and *rubble* – all of which have a similar visual appearance. In particular, the difference between small rocks and large rubble is a matter of opinion, so that the classifier error may not have a major impact on the final biological understanding.

The ICA patches showed markedly better recognition rates over filters, direct patches, and eigenpatches. Examining the results as a function of n and k , shown for both single and multiple models per class in Figure 2-8, indicates that the high rates are not statistical flukes but vary smoothly with the parameters. The variation across the parameter space

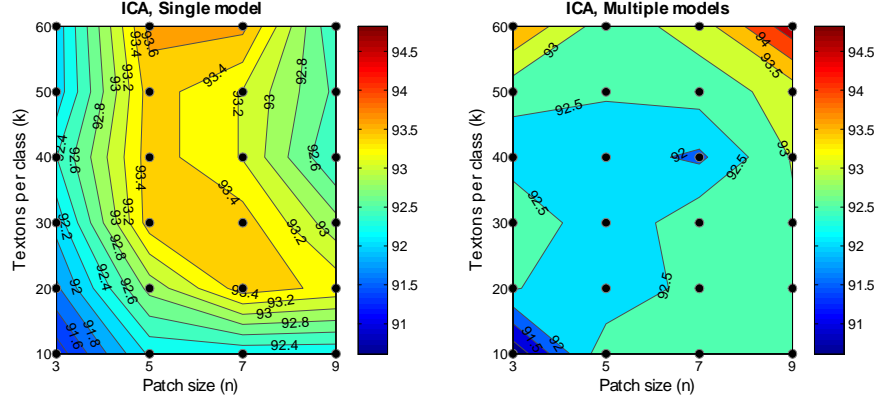


Figure 2-8: Recognition rates as functions of the patch size and number of dictionary textons per class for the ICA transform. The scale is the same for both parameter spaces.

(including models) is only 4.2% – smaller than the difference between means of the methods in Table 2.3. It is interesting to note that the worst recognition occurred for both the single and multiple model cases when $(n, k) = (3, 10)$. The combination of small patch size and small dictionary may not have allowed enough description of the classes to provide adequate discrimination. There is also evidence that both model methods have improved performance for larger dictionaries – but that as the flexibility in describing the model space increases (i.e., increasing from single to multiple models per class), more information is needed in the patch size and number of textons. It is also interesting to note that the single model case has a plateau at $n = 5$ to $n = 7$, suggesting that some minimum patch size is necessary to describe the habitats, but that too large a patch may be too specific and actually decreases the recognition rates.

2.1.5 Habitat area coverage

The motivating goal for creating a habitat recognition engine is to use the classifier to automatically map out environments at dive sites and to estimate area coverage of the different classes. Obtaining a rough estimate of the habitat area coverage would be possible by classifying images spaced far enough that there is no overlap. The SeaBED AUV makes it possible to go one step further and compute a high-quality map of the survey habitats: the images from the example dives were captured at a fast enough rate compared to the AUV velocity and altitude to result in a 20% to 50% overlap between consecutive pictures

(see, e.g., [284]). The degree of overlap makes it possible to align the images to create a continuous map while simultaneously reducing the error by classifying portions of the images in the overlap region two or more times from slightly different viewpoints.

Estimating the area coverage was done in three steps. In the first, each of the images were subdivided into blocks. The best classifier from the previous section, ICA patches-M ($n = 5$, $k = 60$), was used to find the best class estimate for each block; an example is depicted in Figure 2-9. Next, SIFT features [232],[233] were computed on two consecutive images and the two were registered by using RANSAC with an affine transformation between pictures [43],[417],[117]; Figure 2-10 shows an example alignment. The assumption of an affine transformation allows for scale, rotation, and translation between two images [368], reasonable for an AUV maintaining approximately the same or a slowly changing perspective between images. Finally, the overlap between image blocks was calculated using the transforms computed between frames. Figure 2-11 shows an example of calculating the overlap and determining the weighting factor for each block based on the number of pixels it contains which are not represented in any other image and the number of pixels appearing in two or more images. The weights, the AUV altitude, and the classifications were used to compute the area of each class present at each image point. The habitats for the first Daisy Bank dataset is shown in Figure 2-12. The Daisy Bank (dive 3) dataset is strongly homogeneous, containing mostly *mud* in the upper portions and *rocks* in the deeper areas. This is similar to the other dives, where most of the habitat was dominated by one or two types – and usually found within a particular area. This suggests that incorporating Markov chains for fast habitat estimates [353],[106] or Markov random fields [191],[368] for high-resolution mapping would be appropriate for future recognition work.

2.2 Detection of rockfish

The AUV images of the floor used for habitat classification include a variety of species. The interest in this section is detecting and identifying rockfish from those images, a commercially important group of bottom-dwelling fish. There are more than 70 species of rockfish that live off the West Coast in the areas where the AUV source images were captured

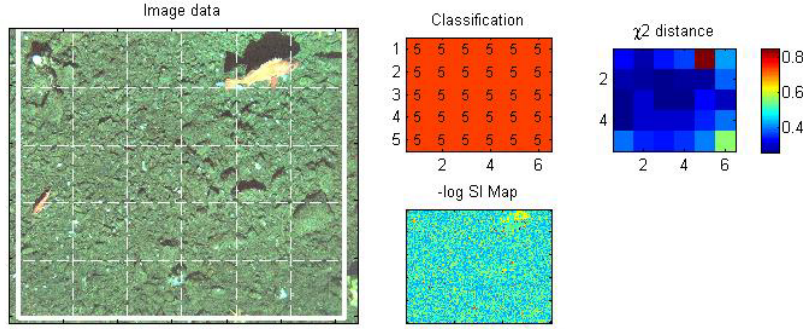


Figure 2-9: Example of classifying an image into habitat blocks. To the left, an image is divided into 24 blocks, denoted by the white lines. The class of each block is estimated and recorded (top center). The top right plot shows the χ^2 distance of each block to the nearest class model. Note that the block with the fish and its shadow has a higher χ^2 distance, but is still recognized correctly. The bottom center plot, $\log(SI)$, serves as an indicator of the frequency of that pixel's textron appearing in the class dictionary and was used as an internal diagnostic.

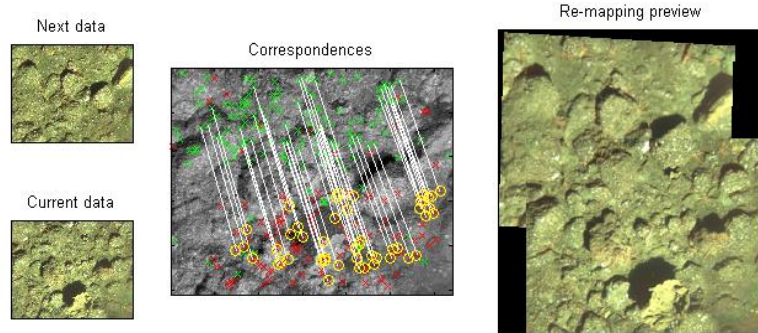


Figure 2-10: Alignment of two consecutive images using SIFT and RANSAC. On the left are the two images which are to be aligned. In the center, green and red dots mark the SIFT features used to compute the alignment; yellow circles denote the corresponding points in the best RANSAC model, with white lines connecting between the dots to indicate the movement between frames. To the right is the result of applying the transform to align and stitch the images together into a single swath.

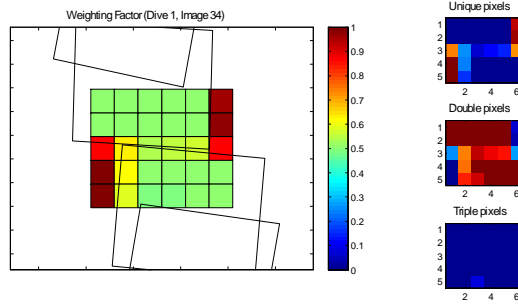


Figure 2-11: Computing the area overlap between several consecutive frames. To the left, the blocks of the current frame are shown with the outlines of the next two frames and the previous two frames overlaid. The color of each block denotes the amount of overlap it shares with other frames. To the right, the weighting factors for each block are broken down into how many unique pixels the block contains, how many pixels appear into two images (“double pixels”), and how many pixels appear in three images (“triple pixels”).

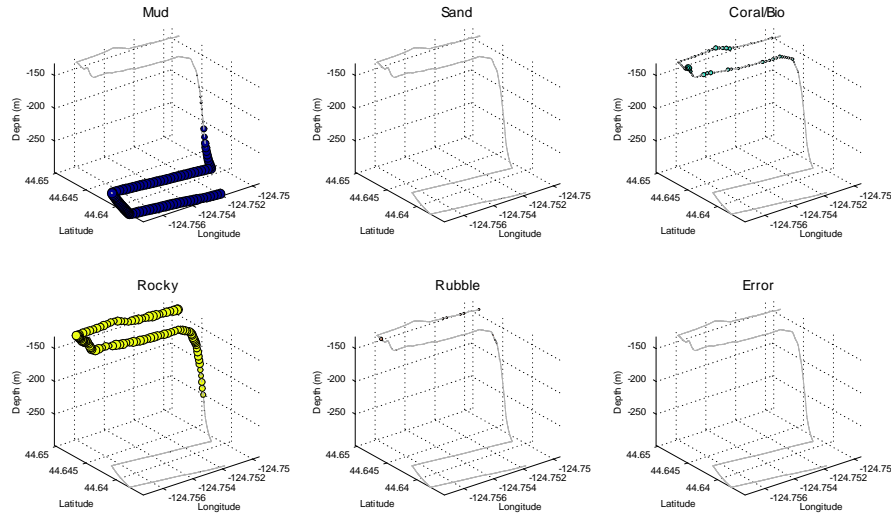


Figure 2-12: Estimated coverage area for each habitat class (Daisy Bank dive 3). The gray line is the AUV’s position over time. The size of each dot indicates the amount of each class present at that location; for visualization, ten images were binned together for each dot present. The dive contained mostly *mud* and *rocks* with limited *coral/bio*, *sand*, and *rubble* with no *errors* detected.

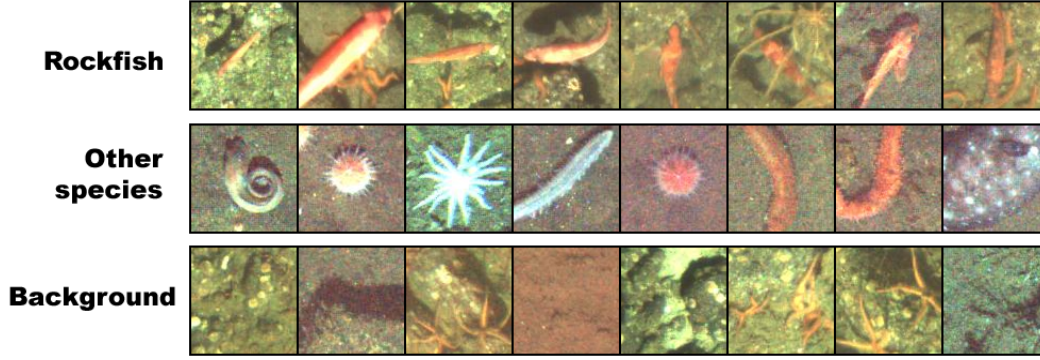


Figure 2-13: Examples of rockfish (top row), other biologicals seen in the dataset (middle row), and backgrounds (bottom row). The backgrounds include biologically active areas where rocks were heavily covered with grasses and brittle stars. Some of the other species present also proved challenging to discriminate against rockfish due to similarities in shapes and colors.

[255],[231]. Rockfish are estimated to have life spans of 100 or more years and mature slowly [231],[51]. Overfishing has had significant impacts on their numbers with warnings about extinction for a number of species [260], leading to emergency closures of many Pacific fisheries areas in the early 2000's [49],[50],[27], with outcry from commercial and sport fishing [242],[26]. Information about the numbers of rockfish, their sizes, and where they can be found is crucial to not only understanding their life cycle and status but for setting appropriate and responsible fishing levels [171],[394],[51],[34],[365],[427].

The goal of this section is to develop a detector which can find images of rockfish in the SeaBED imagery; examples are depicted in Figure 2-13. Results from the detections can be presented to a human expert for time-efficient verification and species-level classification, or used directly with the measured error rates to estimate rockfish counts. Size information and class membership probabilities are by-products of the detector. The detector itself is composed of three stages: the first finds candidate regions, the second uses rudimentary shape information to prune out noise, and the third combines shape and color characteristics to better classify objects as rockfish or non-rockfish. Effort is made to maintain a low removal rate of rockfish at each stage.

2.2.1 Boosted detection

The first step for finding rockfish is to determine all the possible images of fish in the AUV images. Once those candidate image regions have been found, they can be used in later classification stages. The goal for this subsection is then not to necessarily describe the regions and their contents perfectly, but to suggest areas that share many of the same characteristics as the rockfish of interest.

The initial challenge is to separate out foreground objects from varied backgrounds. Controlled environments make detection easy by artificially creating a background which has statistics vastly different from the objects of interest so that regions of interest can be determined by thresholding or a simple statistical analysis [419],[361],[311],[442],[79]. The SeaBED images have distinctly non-trivial backgrounds: clutter, texture, color, lighting, and competing species all make the background hard to predict – even with the habitat detection methods developed in the previous section. For example, one immediate idea is to mark areas with statistically improbable textons (given the region’s habitat classification) as objects. The problem with this is two-fold: there is enough natural variation that statistically improbable regions are not reliable indicators of foreground object regions, and similar textons as used for the backgrounds are present on the objects of interest. One good example is textons generated from biologically active rocky areas, where the natural flora is a good indicator of the habitat type and shares characteristics (such as color) with rockfish. The end result is that statistical thresholding, either on image features or textures, does not provide a working solution.

The alternative taken here is to create a boosted classifier that uses multiple features to indicate whether a pixel should be foreground or background, providing the same result as a detector. Boosting is a flexible meta-classifier that has the ability to combine a series of weak classifiers into a stronger classifier. For images, thresholding on a single feature provides a simple and direct weak classifier, and is suitable for boosting if the thresholding is able to correctly classify at least better than randomly. The decision boundary created by a series of boosted thresholds need not be linear, either, able to construct complicated boundaries through the feature space without necessarily needing a kernel transform or prior knowledge of the distributions of the features [127],[129],[269],[368].

A selection of color and structural image features were used for the first round of boosting. A training set of 410 rockfish and 1881 non-rockfish patches (negative detection samples; 1502 samples of ground/texture and 379 samples of species other than rockfish) were randomly selected and extracted from the SeaBED images of the previous section and used to compute the training features. A hue-saturation-value (HSV) color space was found to be more discriminative than an RGB color space as most rockfish have a high saturation and value components while many of the naturally occurring backgrounds have lower saturation. The mean and standard deviation of each HSV color component was calculated in a local neighborhood. Since the hue value is an angle, statistics of the cosine and sine of the hue were used instead. Structural information was provided by a series of monogenic filters at four scales (bandpass regions) computed on saturation and value component images [114],[198]. The mean energy and standard deviation of the filter responses was retained. A local neighborhood around each pixel of interest was computed by using Gaussian weights for the local mean and standard deviation. A spatial variance of $\sigma = 5$ was found to be better than larger neighborhoods.

Training performance of the boosted classifier using LogitBoosting and the HSV-based features is shown in Figure 2-14. The boosting is repeated twenty times with a random 20% of the samples removed for testing, giving rise to the depicted uncertainty at each stage. The learning plots indicate excellent detection for the training samples after only a few rounds of boosting; each round corresponds to adding a weak classifier with a threshold selected to minimize the weighted training error [127],[129]. Figure 2-15 shows examples of applying the boosted detector to each pixel in an image. (See Section 4.1.3 for more discussion on using boosted decision stumps as a detection filter in image processing applications.)

The actual performance of the boosted detector is slightly lower after including masking operations (morphological filters to remove detections considered to be too small and connecting nearby detections) and testing on complete images, with around 7% of rockfish going undetected through the entire dataset. The detection rate is strongly dependent on the habitat type, Table 2.5, where the habitat was determined by the dominant environment using the methods of Section 2.1. The table represents 10% of the images from the data set, randomly selected for hand verification. Areas with rocks and higher numbers of other

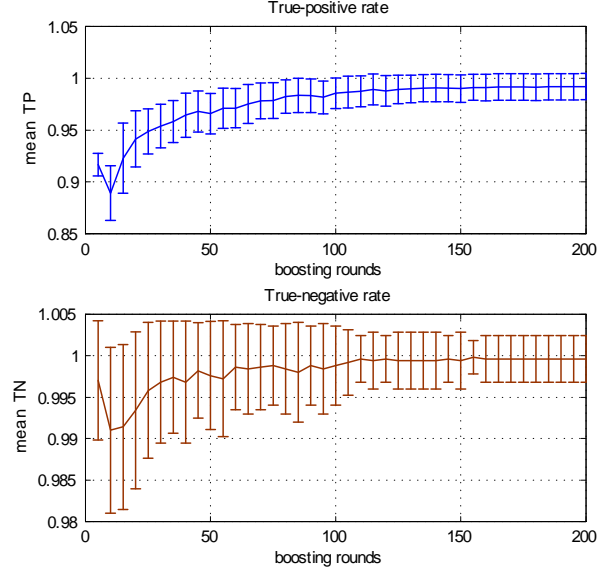


Figure 2-14: Training rates for initial object detection using boosting. Error bars are $\pm 1\sigma$, computed by repeating the training twenty times while holding a fraction of the data for testing.

species led to the lowest detection rates due to occlusions, shadows, and other species with similar colors and shape such as brittle stars. Fortunately, this suggests that the true counts can be estimated based on the habitat and its known error rate [347],[165],[166],[423].

2.2.2 Shape pre-processing and spurious detection removal

The first stage located regions that shared characteristics with samples of rockfish based on their color and basic structure. For the example dataset, it returns more than 130,000 detections for around 30,000 source images, an average of 4.3 regions per image. Reducing the number of detections by starting to analyze the contents of the regions is the goal of the second stage – and also the beginning of a transition between detection and classification. The two share a particularly fine line at this juncture because the detection process already includes some classification decisions related to size (i.e., rockfish are assumed to have some minimum size during mask creation), and the second stage uses shape estimates which could be considered size measurements in their own right.

Computing features on an object again means that the background and the foreground

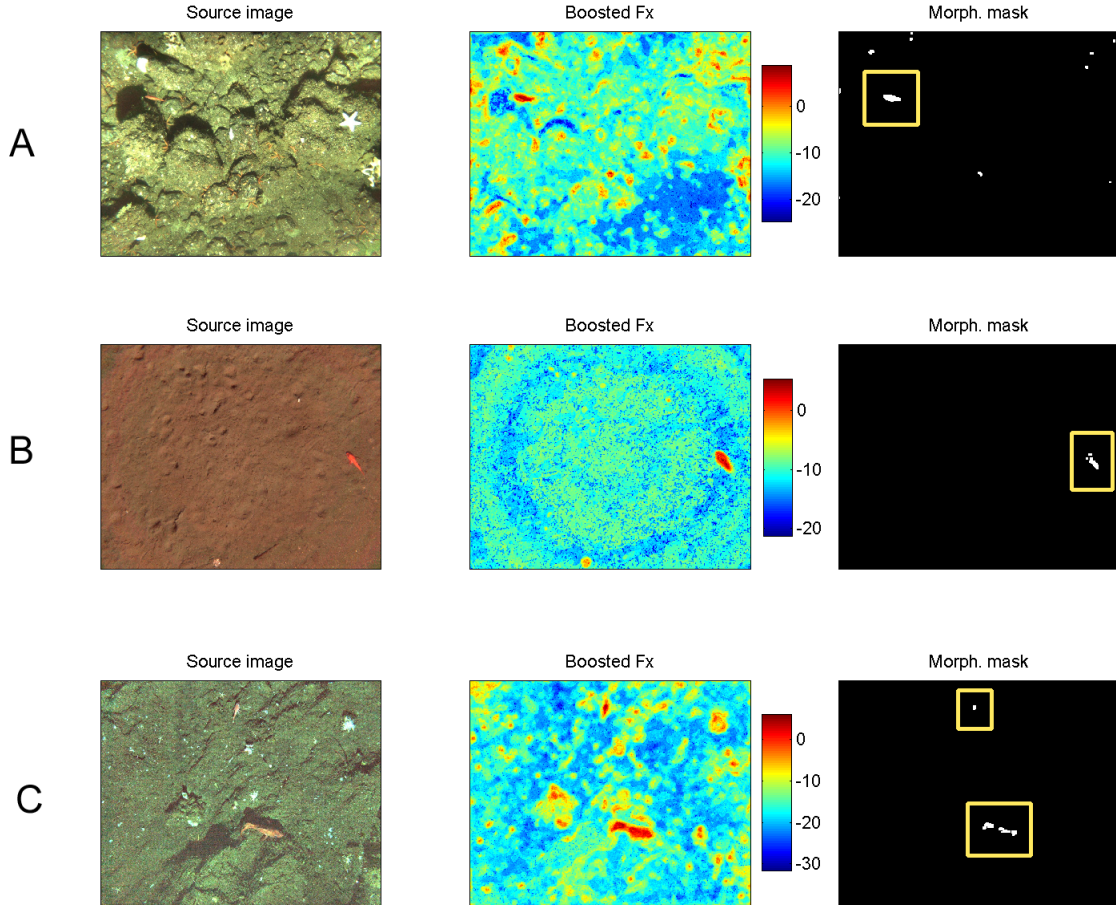


Figure 2-15: Boosting using HSV-based color and structure features selected for rockfish. Left column: sample images from three different habitat types. Center: pixel-wise classifier response at each pixel; a positive Fx is a strong indicator that the area has rockfish-like features, while a negative response indicates background. Right: a mask constructed from setting $Fx \geq 0$ and applying morphological operators. Yellow boxes surround correct detections of the rockfish in the images. Note that sample A contains a number of other bright red features, some of which are detected and will need to be filtered in later stages. The mask from sample C is disjointed, but will be corrected during the second classifier stage.

Habitat	Images	Rockfish det.	True count	Det. rate	#/image
Mud	454	78	79	98.7%	0.17
Sand	1848	141	145	97.2%	0.08
Coral/Bio	26	32	33	97.0%	1.27
Rocks	420	560	618	90.6%	1.47
Rubble	220	81	83	97.6%	0.38
Overall	2968	892	958	93.1%	0.32

Table 2.5: Detection rate of rockfish as a function of the habitat type for a randomly-selected subset of images. Rock and coral/bio habitats have a larger number of occlusions, confounding species, texture variation, and shadows. The higher prevalence of nooks and crevices may also lead to more juvenile or smaller rockfish in those areas (consistent with Tolimieri et al.), also contributing to more challenging detections. The average number of rockfish per image is based on the hand-verified true count over the sample.

need to be separated. The previous discussion about detection provided some estimates about a local region on a pixel-wise basis. Thresholding and morphological operations on the boosted detection strengths gives an initial guess for size. However, since the pixel-wise boosting was derived from local area operators, the resulting detection map is smooth and each pixel has the potential for representing information outside of the object. Initial attempts to segment the image regions around the boosted detection blobs using grow-cut caused many of the small areas to disappear completely or the large objects to grow beyond their boundaries. Grow-cut expects a good initialization to determine properties of the foreground and background pixels with high certainty, filling in the uncertain pixels based on its learned properties and the relative pixel positions [401]. The non-locality and smoothness of the boosting-based initial guess was not sufficient.

A level-set segmentation proposed by Chan and Vese [57] is less sensitive to initialization [345] and was found to have excellent results for the rockfish. The Chan-Vese model finds a contour, C , which minimizes the squared error of some value, v , inside and outside a contour while simultaneously minimizing the contour length. For segmentation, intensity or color is often chosen as the value of interest. The idea of squared error implies that a correct value is known; Chan-Vese estimate the correct level with the mean value inside and outside the

contour at each iteration. The minimization functional is then

$$CV(C) = \alpha L(C) + \beta \int_{in(C)} |v - \mu(v_{in})|^2 dx dy + \gamma \int_{out(C)} |v - \mu(v_{out})|^2 dx dy, \quad (2.2)$$

where $L(C)$ is an energy functional on the length or shape of the contour (which can be adjusted for different applications), $\mu(v_{in})$ is the mean of the value inside the contour C , $\mu(v_{out})$ is similarly the mean of the value outside the contour, and the shorthand $in(C)$ and $out(C)$ indicates regions inside the contour and outside the contour. The scalars α , β , and γ adjust the relative importance or weighting for each term. Additional energy terms can be included as needed [57]. It is interesting to note that the error terms are thinly-disguised variance estimates: normalization by the area or number of pixels can be absorbed into the weighting constants and becomes unnecessary if $in(C)$ and $out(C)$ are nearly equal in size. Equation 2.2 can then be thought of as finding a segmentation that minimizes the variance or that maximizes the consistency within each group.

A consistent value for the segmentation was available from the earlier HSV-based metrics. The rockfish tended to have high color saturation compared to their surroundings and similarly had a higher brightness. The product of the saturation and value channels provided a reasonable key for discriminating objects. Fast shape metrics are computed on the Chan-Vese regions masks, including area, perimeter, and the size of a fitted ellipse (major axis length, minor axis length, and eccentricity).

Two different methods for quickly pruning non-rockfish objects were examined. First, a linear discriminant analysis (LDA) was tried for the sake of providing a baseline. The LDA seeks to maximally separate two classes by finding an optimal projection direction (see Section 2.3.1 and [106] for background and discussion). The optimality assumes multivariate Gaussian distributions for the data in each class, which is decidedly not the case for positive-valued size descriptors. A log transform increases the Gaussianity of the distributions enough so that, while not guaranteed to be optimal, the discriminability of the LDA projection is significantly increased.

A test set of 500 rockfish regions and 2500 non-rockfish regions were used to estimate removal rates for different combinations of features. The LDA was computed for each set of

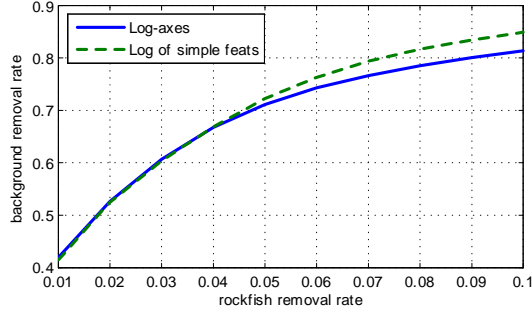


Figure 2-16: Removal rate of background objects based on a threshold chosen to limit the expected removal rate of rockfish regions. The dotted green line is computed with the log-transform of all five simple shape features, while the solid blue line is the removal rate achieved with only the log transform of the major and minor axes of a fitted ellipse. At a 2% predicted rockfish removal rate, more than half the non-rockfish detections are removed; a threshold on the LDA-projected value which results in a 5% rockfish removal rate leads to around 70% removal of non-rockfish detections.

features and the samples of each class were projected onto the resulting LDA direction. A kernel density estimator (KDE) was used to compute the cumulative distribution functions (CDF) of the projection for each class. Rockfish and background object removal rates are plotted in Figure 2-16 for the entire set of log-transformed size variables and for only the log of the major and minor axes. The fact that the log-transformed axes have a similar expected non-rockfish removal rate suggests that the majority of the discriminative information lies in those two features. Indeed, the other features are closely related to the log-transformed axes lengths: the log-area is proportional to the sum of the log-axes, and the log-eccentricity can be shown to be approximately linear with the log-axes for eccentricities close to unity (as would be the case for rockfish; the monogenic filters also biased the detections towards areas with higher eccentricities). Perimeter is closely related to area (see [57] and their reference 7), so that much of that feature is contained in the axes features as well.

The second rockfish-background discrimination method again uses boosting, this time to specifically create a non-linear decision boundary and with a modification for the desired loss rate. To see how the boosting should be modified, it will help to understand how the classifier is created under normal use where there is no specific loss function. Adaptive boosting (AdaBoost) creates a strong classifier in an attempt to minimize an indicator

function of the form

$$J(F(x)) = \left\langle e^{-CF(x)} \right\rangle, \quad (2.3)$$

where $C \in [-1, 1]$ is the true class label⁶, $F(x)$ is the classifier indicator function constructed on the features of sample x and a series of weak classifiers, and $\langle \cdot \rangle$ denotes taking an expectation. Then

$$\left\langle e^{-CF(x)} \right\rangle = P(C = 1|x) e^{-F(x)} + P(C = -1|x) e^{F(x)}, \quad (2.4)$$

and by differentiating with respect to $F(x)$ and solving for the zero, the optimal $F(x)$ satisfies

$$F(x) = \frac{1}{2} \log \left(\frac{P(C = 1|x)}{P(C = -1|x)} \right). \quad (2.5)$$

Classification is done on the $F(x)$ response by setting $C = 1$ for $F(x) \geq 0$ and $C = -1$ for $F(x) < 0$. This is then selecting for the maximum class membership probability based on the training data [129]. Incorporating a risk, such as for miss-classifying a rockfish as a background object, is as simple as adding weights (or priors, for that matter) to the selection,

$$\begin{aligned} F_\lambda(x) &= \frac{1}{2} \log \left(\frac{\lambda_1 P(C = 1|x)}{\lambda_{-1} P(C = -1|x)} \right), \\ F_\lambda(x) &= F(x) - \frac{1}{2} \log \left(\frac{\lambda_1}{\lambda_{-1}} \right), \end{aligned} \quad (2.6)$$

where λ_1 and λ_{-1} serve as the risk for the $C = 1$ and $C = -1$ classes. The risk terms can be used as additional weighting factors in Equation 2.3 to specifically generate a new F_λ , or as Equation 2.6 shows, simply used to shift the decision threshold of the original $F(x)$ classifier constructed without weights. This second, more intuitive alternative is preferable since it requires little a priori knowledge about a correct λ ratio during the training phase and the threshold can be adjusted after training.

⁶The peculiar restriction of $C \in [-1, 1]$ for two-class AdaBoost actually serves as an indicator: $F(x)$ is designed to have the same sign as the sample's true class, and CF is then positive when the classifier and the training data agree or negative when they differ. The amount of misclassification, a larger $|F|$ with an erroneous sign, leads to larger exponential weighting.

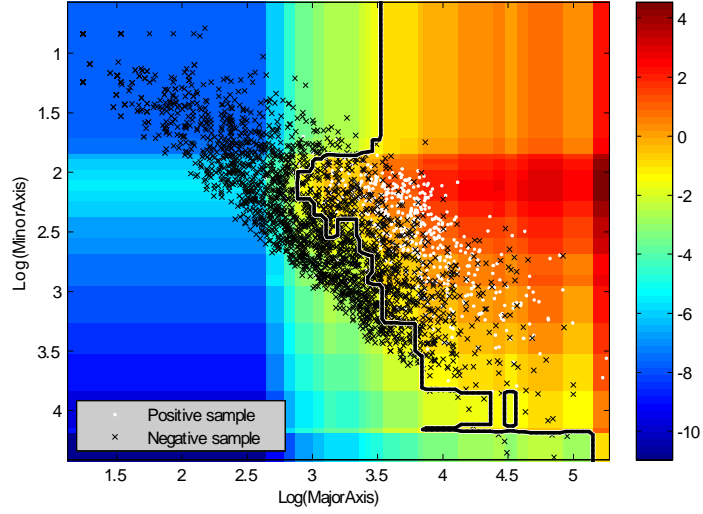


Figure 2-17: Decision landscape for a boosted classifier trained to lose 2% of rockfish samples based on simple shape descriptors (the same log-axes features as the LDA of Figure 2-16). The thick line indicates a decision boundary on $F(x)$ with the actual $F(x)$ value depicted as the background color. Training samples are marked as white dots (positive: rockfish) and black \times 's (negative: background). The boosting uses decision stumps, which lead to rectangular decision areas.

A classifier was created using LogitBoosting with 80 stump classifiers and the log-axes features on the same data set as used for the LDA tests. A CDF was estimated with a KDE and used to find a threshold for $F(x)$ corresponding to a 2% rockfish removal rate for the training data. The threshold corresponds to a 72% background removal rate, around 20% higher than for the LDA at the same expected rockfish removal rate. Figure 2-17 shows why the boosting has better performance. The LDA uses a single decision threshold in the projected data, corresponding to dividing the feature space into two half-spaces. The boosted decision threshold is actually a complicated line, going nearly diagonally through the feature space, then transitioning to a nearly vertical cut-off around $\log(a) = 3.5$ (where a is the major axis length in pixels). The negative sample points in the upper left of the feature space are specifically excluded by the boosting.

Figure 2-16 shows two other important aspects of boosting. The rectangular shaped areas are a result of using decision stumps as weak classifiers. Each stump is built on a single feature and threshold. The diagonal decision boundary suggests that, while enough

decision stumps can reproduce complicated boundaries, linear classifiers constructed on both variables would have been better choices for this case (especially since processing time is not an issue with this small of a number of samples and features). There is some evidence of overfitting as the decision threshold zig-zags around regions to fit data points. The second point is that enough data is necessary to shape a complicated decision boundary. Here, the limited number of positive samples does not adequately flush out the diagonal decision line.

Checking the performance of the LDA and boosted classifiers for removing spurious background was done by resampling the training data and applying the classifiers to the new samples. A Gaussian mixture model (GMM) was fit to the positive and negative training data, resulting in two mixture components for the rockfish samples and four mixture components for the background. The higher number of mixture components for each class guarantees that the LDA is not optimal since each class is not fit well with a single Gaussian. The non-Gaussianity also suggests that results from resampled data will not be identical to results from the training data. In fact, the LDA classifier removes 1.3% of the rockfish while discarding 53% of the background. The boosted classifier removes 4.4% of the resampled rockfish and 71% of the background. The slightly higher removal rate from the boosted classifier may be an effect of overfitting to the training data.

Finally, it is worth noting that both the LDA and boosted classifier provide simple metrics for predicting a sample’s membership in either the rockfish or background class by estimating the distributions of the LDA projection or the $F(x)$ values. These probabilities can be combined with later stages rank samples and present only those with the greatest uncertainty to a user for verification.

2.2.3 Shape-based classification

The second stage selected against non-rockfish objects which could be quickly identified using simple size features and fast calculations. The remaining detections require more in-depth features.

Sirovich and Kirby proposed the use of an eigenvector decomposition for face matching in a representation now known as “eigenfaces”. A series of sample face images are sized and aligned, then each face is reshaped into the column vector, providing the columns in a data

matrix. The principle components of the data provide a compact basis that is optimized to represent the variance. Matching is done by finding the shortest Euclidean distance (or other preferred distance metric) between the vector of basis coefficients of an unknown sample and the known database of faces [339],[192],[388],[24],[434].

A similar idea was tested with the fish shapes. The Chan-Vese detection regions were resampled to a fixed number of pixels and aligned using the orientation of the major axis of the fitted ellipse. A singular value decomposition of the aligned region information gave a compact basis for describing the shape. Four bases were used for experimentation, one for the rockfish regions, another for the background objects, and two bases computed on the region outlines of the training set rockfish and non-rockfish. Initial experiments using the bases for classification showed significantly better results from the outline bases over the region bases. A resampling size of 40×40 pixels gave the best classification for the training samples, balancing computation time, specificity, and dimensionality.

Example “eigenfish” from the positive (rockfish) and negative (background) samples are shown in Figures 2-18 and 2-19. The first few eigenfish (or, more appropriately, “eigen-nonfish”, for the negative samples) corresponding to the largest singular values are bar-like and edge-like, and describe how the energy is distributed around the axis in more detail than the major and minor axes values of the second stage classifier. The P_4 and P_5 eigenfish indicate a narrower tail than the body, with an ambiguity in the orientation direction. The highest-energy positive and negative eigenfish share a large number of characteristics, which indicates that they might not be particularly discriminating for classification. Eigenfish selected by a best-first heuristic [152],[106] are shown in Figure 2-19 in order of selection. Each of these eigenfish show greater amounts of asymmetry, suggesting that the variations are actually more useful for classification.

The shapes in Figure 2-19 are also reminiscent of Zernike polynomials (see, e.g., Figure 4-11). Zernikes and other image moments have a long history of use for image recognition [188],[63],[25],[183],[372]). Normalized Zernike moments were calculated on the same regions and outlines without resampling and realignment as another shape feature.

Classification tests on the eigenfish and Zernike moment features were run ten times each with 66% of the training data randomly selected for training and the remaining samples

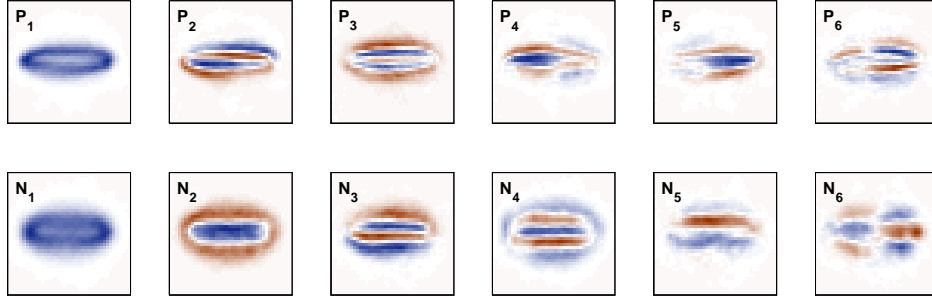


Figure 2-18: The first six eigenfish outlines in order of the energy each represents. The top row is the eigenvectors of the positive training samples (rockfish). Negative training samples (background) are on the bottom row. Red and blue colors indicate opposite signs, with white being zero.

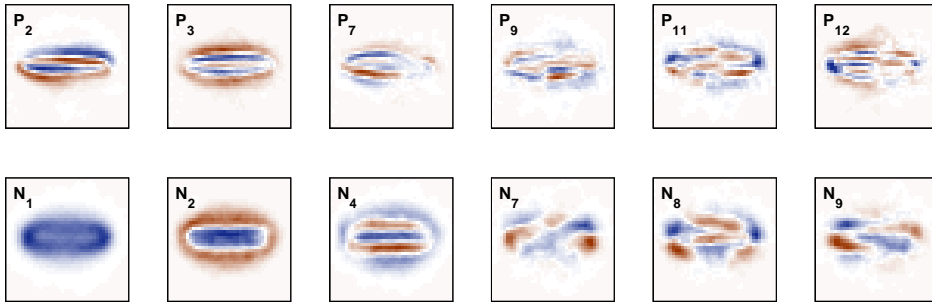


Figure 2-19: Eigenfish outlines selected by Weka as potentially useful features for classification tasks, in order of their ranking. The number indicates which eigenvalue the image is derived from, with the greatest energy in the smaller eigenvalue indices. Positive (rockfish) samples are on the top row and negative (background) samples are in the bottom row. Red and blue indicate opposite signs with zero as white.

used for testing. A naïve Bayes classifier achieved around 78% correct classification for both Zernike and eigenfish feature sets and provides a comparison baseline. Trees and different boosting schemes are able to achieve around 79-81% overall correct classification for the Zernike moments; boosting gets 79% correct for the eigenfish while trees are able to achieve a 74% rate. A k -nearest neighbor classification was also used with $k = 5$ neighbors, following the pattern matching methods applied with earlier eigenface work, achieving 72% and 81% for eigenfish and Zernike moments respectively. In general, the Zernike moments provide better classification with much less computation and without needing a good orientation estimate.

The first detection stage incorporated information about local color and its spatial distribution. Detections were counted as areas with enough boosted pixels that were above a fixed threshold, a limited measure of the distribution of the first-stage boosted detector response $F_1(x)$. By considering each pixel's response as a weighted vote for a class membership, the full distribution of $F_1(x)$ within the object region can be thought of as a population weighted vote for the object. The cumulative distribution functions (CDF) for rockfish and background samples which passed the second-stage boosted classifier are plotted in Figure 2-20, with the rockfish CDFs lying to the right of the background object CDFs (and with better separability near the high $F_1(x)$ region). The visible separation between the two sets of CDFs suggests an additional set of features: the $F_1(x)$ values where the CDF reaches prescribed probability values. This is a discrete representation of the inverse CDF and provides a description of both the shape of the curve and its position. The set of inverse CDF features gives around 79% overall classification accuracy using 80 rounds of LogitBoosting, with indications that improvements are possible with additional boosting rounds.

Classifiers using combinations of the Zernike-based shape descriptors, detection inverse CDF values, and fast shape metrics (eccentricity, perimeter, circularity, and perimeter-to-area ratio) were tested with LogitBoost and eighty decision stumps. The results were averaged across ten runs of a leave- n -out scheme, where 66% of the data was randomly selected for training and the remainder held for testing in each iteration. Overall classification rates, Table 2.6, suggest that the inclusion of all three feature sets is worthwhile, especially

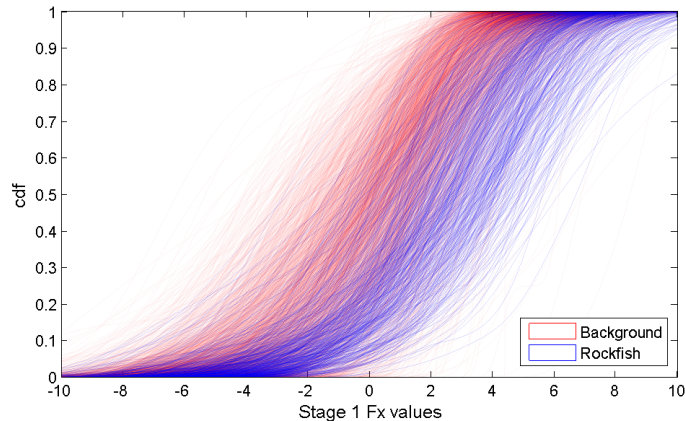


Figure 2-20: Distributions of the boosted detector response within Chan-Vese object regions of training objects which passed the second-stage classifier. The CDF curves of the rockfish samples (in blue) lie to the right of the non-rockfish samples (in red), indicating a consistently higher response strength through much of the region aside from the distributions tails. The distributions overlap more at the bottom (the area of negative $F(x)$ and low CDF) than at the top of the curves.

since all three sets are fast to compute.

Multi-stage classifier in use

A classifier with 400 decision stumps and LogitBoosting was built with all three feature sets for the third stage of selection. Testing was done using hand-verified detections from 10% of the images (the same as used for Table 2.5), a total of 892 rockfish and 10,482 non-rockfish; 845 rockfish and 3346 non-rockfish detections survived the boosted second stage classifier. The third stage had an 88.6% overall classification accuracy, 80.4% accurate with

Zernike	$F(x)$ ICDF	Fast Shape	Est. TP Rate
×			79.8%
	×		79.1%
		×	83.0%
×	×		84.6%
×		×	84.7%
	×	×	85.6%
×	×	×	87.0%

Table 2.6: Initial tests on combinations of three feature sets for the Chan-Vese regions. An x marks which sets were used in that particular test. The estimated overall true positive rate is averaged over ten runs.

Habitat	Images	#Rockfish	Stage 1	Stage 2	Stage 3	Det. rate	Prec.
Mud	454	79	78	74	57	72.2%	0.75
Sand	1848	145	141	138	121	83.4%	0.65
Coral/Bio	26	33	32	28	12	36.4%	0.25
Rocks	420	618	560	526	419	67.8%	0.71
Rubble	220	83	81	79	70	84.3%	0.81
Overall	2968	958	892	845	679	70.9%	0.69

Table 2.7: Numbers of rockfish for each habitat through the detection-classification stages. The data comes from a hand-verified random sample of the original image set.

rockfish and 90.7% with non-rockfish. The precision for returning rockfish was 0.69, so that the number of false positive from background objects is still high but not nearly as dominating as after the first stage of detections (a precision of 0.085). Results for each habitat type are summarized in Table 2.7. The two habitats with the lowest detection rates have the highest number of occlusions and confounding species (in particular, brittle stars). Fortunately, the coral/bio class has few images throughout the dataset, and the detection errors are less detrimental to the overall counts.

Boosted classifiers are expected to have high response strength (high $F(x)$) for objects that have a high probability of belonging to the positive and negative sample classes. A portion of the test samples classified as rockfish after the third stage are shown in Figure 2-21 in order of decreasing classification strength from left to right. The majority of the errors are in the right side of the image, corresponding to low classification strength, as expected. This is also the area where a human observer can make the most of their time, verifying the detections which were particularly challenging for the computer.

Great care is taken to ensure that measurement devices have a negligibly low bias or a constant bias which can be reliably corrected. The SeaBED AUV is a particularly interesting case in that the optical design can be characterized during the design phase but the biological component of bias can only be assumed or estimated. The SeaBED has a potential issue with avoidance, scaring away target species with bright strobes [384]. There is approximately a one-third overlap between consecutive frames with the top portion of the n^{th} image aligning with the bottom portion of the $(n + 1)^{th}$ frame (Figure 2-11 shows an example of the overlap). If the SeaBED strobes are causing rockfish to hide or flee, this

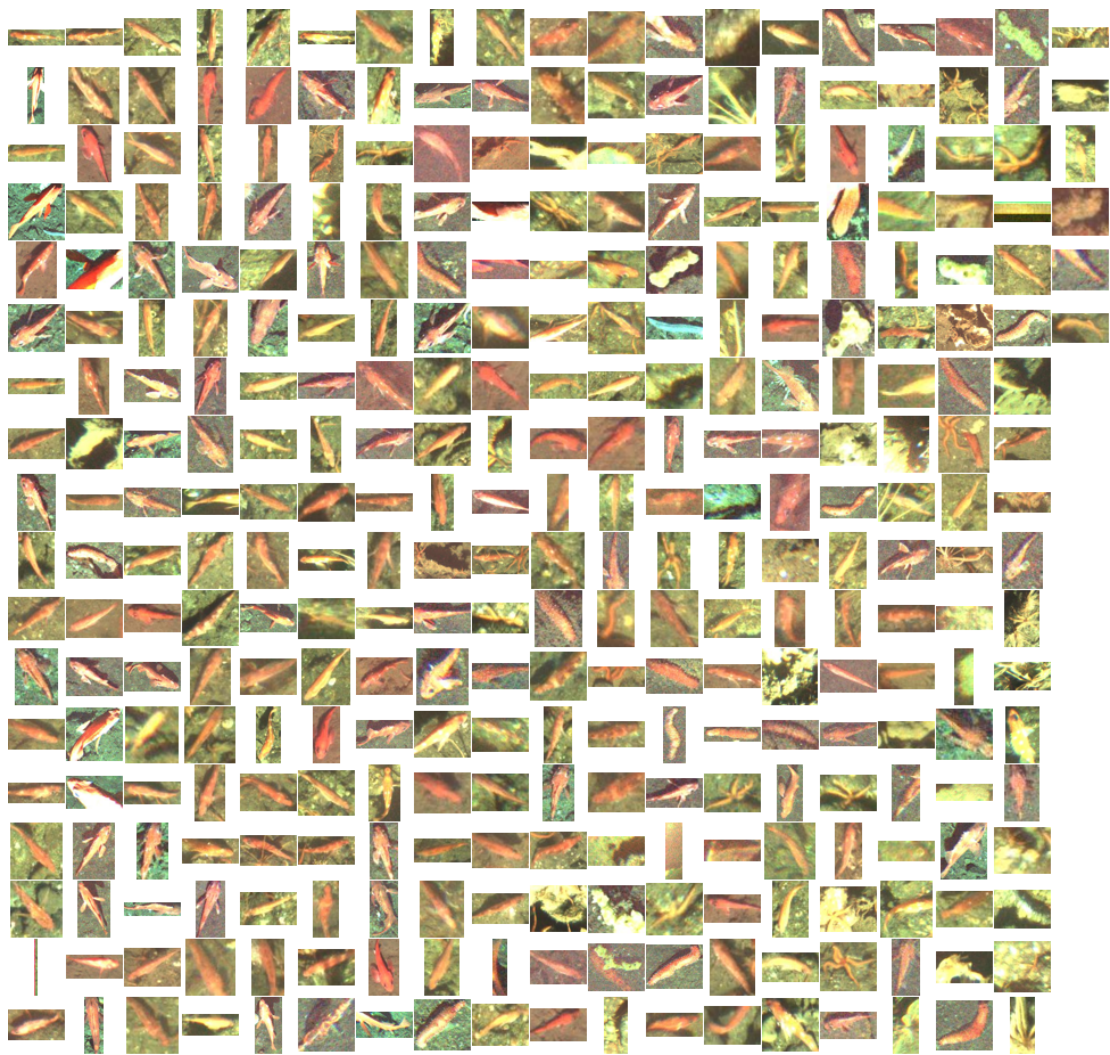


Figure 2-21: A selection of objects returned by the multistage detector-classifier as possible rockfish. Samples are arranged in order of decreasing third-stage classification strength from left to right (column-major order). Most non-rockfish objects are found in the right side of the collage, corresponding to a low class membership probability.

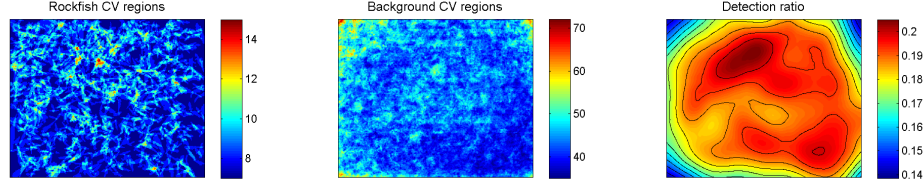


Figure 2-22: Left and center: spatial distributions of rockfish and background object regions from the hand verified data. Right: detection ratio of rockfish compared to non-rockfish objects, averaged over a local area.

would appear as a lower detection rate for rockfish in the lower part of the images.

The hand-verified data used for testing during the classification stage sheds some proverbial light on avoidance. The Chan-Vese regions for both known rockfish and known background objects are superimposed together in Figure 2-22. The total number of detections is higher towards the top half of the frame, which may be partly due to the angle of the strobes relative to the frame and stronger shadows to differentiate the structure and shape of objects. The key, however, is that the ratio of rockfish detections to non-rockfish detections is nearly constant over the frame, 0.18 - 0.21 rockfish detections to non-rockfish detections per pixel on average. The edges and corners have greater amounts of noise and erroneous detections, decreasing the detection ratio in those areas. Constancy over the majority of the frame strongly suggests that the rockfish are not preferentially avoiding the SeaBED unit.

The rockfish and habitats share a similar problem of over counting. The image capture rate of the SeaBED means that consecutive images have overlapping areas, and rockfish in those regions can be counted two or more times. The SIFT-RANSAC image registration discussed in Section 2.1.5 for habitats can also be re-applied here to determine which detections might match between frames. Orientation estimates of the rockfish provide clues for the direction of travel so that motion may be easier to accommodate. Overlap in this case may actually be beneficial if distributions for size or swimming speed are known for the rockfish: the length and velocity (both in pixel units) might help estimate the height of the rockfish above the seafloor and thus correct the size for camera perspective. Similarly, shadows can be used in relatively flat habitats (e.g., mud, sand, and rubble) to estimate the

height of a rockfish above the floor. These ideas are left as future work, all made possible by a reliable detection process.

2.3 Longline detection and linking

Longline fishing refers to the practice of fishing by distributing hooks and bait along a trawl line several hundred meters to kilometers long. The longline may be weighted so that it sinks to the ocean floor (bottom or demersel longline; associated with halibut, sablefish, rockfish, and groundfish) or suspended in the water column (pelagic longline; associated with swordfish and tuna). In Canada, longlines account for around 6% of the commercial catch by value and about 3-4% by weight [132]. The US is similar. The line is pulled back into the fishing vessel and the catch is immediately pulled from the hooks as it comes into the boat. Monitoring and tracking the number of each species caught on the line is currently done by a trained observer, either on-board the vessel or by reviewing video footage of the operation afterwards, and is only done for a fraction of the fishing fleet. Even reviewing the video afterwards can be time consuming, especially given the number of boats actively fishing any one area – so that critical decisions from the fisheries management can be delayed, leading to potentially dangerous results for the ecosystems and sustainability of the fisheries [171],[430],[251].

Longline fishing has received vocal criticism for its problems with unintended bycatch. Seabirds get caught by the hooks as they attempt to eat the bait or small fish caught by the line [41],[181]. A number of small sharks and sea turtles are susceptible to similar problems [124],[446]. Counts of species unintentionally caught on the lines can help both activists and the fisheries improve practices [132], better protect the ocean environment [430],[171],[124],[427], and even estimate the populations of species which are otherwise difficult to count [258].

The application goal of this section is to use intelligent image processing methods to detect and localize images of fish as recorded by the video equipment on long line fishing vessels, linking together the various views of an animal so that an observer can quickly scan through and count the fish. Along the way, methods are developed for background

removal which is tolerant to motion and for determining which samples are likely the same animal. The final stage of linking samples is particularly challenging because of the low temporal sampling of the video, allowing significant motion between frames and making traditional methods of motion tracking, area deformation [443], and color tracking [280] moot. Priors are limited to thresholds that the user can set interactively, or those priors which the methods estimate and learn.

The sample data for this section comes from real-world longline fishing operations. A webcam with low spatial and temporal resolution was mounted rigidly on each boat using an arm facing towards the entrance chute. Images of 480×640 pixels were recorded at approximately five frames per second and compressed to save disk space. (The examples shown came from a video recorded at 5 fps; the methods here were developed on samples ranging from 5 to 10 fps.) The videos include motion from the water, workers, and the fishing line itself. Typical lengths range from 30,000 frames to 60,000 frames, with the active line retrieval limited to the middle 50-80% of the video. The workers routinely reach down to grab the fish using a hook, tossing back unsalable fish and sharks. Nothing is modified on the boats or the imaging to specifically aid in localization. The line routinely pauses while workers deal with difficult fish or problems in the line itself. An example frame and captures of fish are shown in Figure 2-23.

2.3.1 Detection of anomalous areas

The first step of detecting objects in the video is to determine what counts as an anomaly and what is expected. To that end, estimating the background and regularly occurring objects gives the normal view against which to measure motion. A stationary viewpoint and focus for the camera helps limit the variation. However, slow changes in the illumination over time mean that the background cannot be approximated as a constant. Using a moving average over a number of neighboring frames to estimate the background allows for adaptation to slowly changing illumination and shadows, but leads to false positives from fast-moving structures such as wave occlusions. Weighting the background map by the inverse of the variance creates a map of normalized z -scores and reduces false positives from areas which see higher variance but is still susceptible to large perturbations.



Figure 2-23: Examples from longline videos. Left: looking towards the *Ocean Foam* deck. Fish are pulled up onto the white platform with color stripes; a worker (in purple gloves) is visible above the platform, and the water is visible as a dark area to the bottom of the frame. Right: samples of fish images from the *Southern Dawn*; similar species are visible in the *Ocean Foam* data used as an example in this section. The images have not been resampled or scaled.

The longline videos show a number of regularly occurring structures: occlusions on the side of the boat due to wave motion or global illumination changes from cloud cover and sun direction. This suggests finding representations of these regular structures that can be used to selectively remove background elements [410],[309]. A random sample of N_e frames is taken from the video, median filtered to remove small specular highlights from wave reflections, then low-pass filtered and reduced in size to $M_r \times N_r$ pixels. Resampling the image balances the amount of computation required against the available information contained in the frames, as significant compression and poor optical clarity of the cameras limits the actual resolution and information. Each reduced frame is normalized against its mean to remove global illumination effects, then reshaped into a column vector as a single background sample of size $3M_rN_r \times 1 = N_{p,r}$, where the third dimension comes from the three RGB color components. The vectors are collected into a single matrix of background samples, \mathbf{E} , and the mean value subtracted from each pixel. An implicit assumption is made that the randomly selected frames used to form the background data matrix have a similar appearance as during the time period when the line is retrieved. Utilizing time periods outside of the active time retrieval time helps weight the background estimation

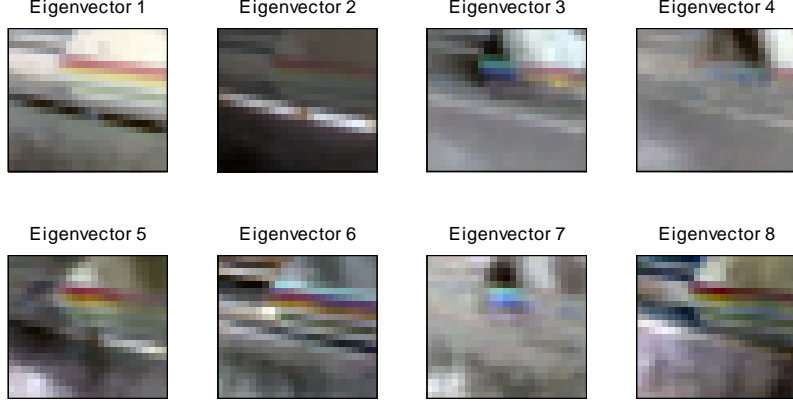


Figure 2-24: First eight eigenvalues of \mathbf{U} for a sample set of frames, reshaped from a vector into an image. The eigenvalues are in order of decreasing energy (singular values). The images are of the side of a boat: the dark area at the bottom is water, the colorful stripes in the center-right are the edge of the platform that the longline is pulled onto, and the white and black stripes to the upper left are the side of the boat. The first and second eigenvectors are due to vertical wave motion that obscures a portion of the hull. The effect of shadows and splashing can be seen in eigenvectors 5 and 8 especially.

towards regular structures which are not related to the fish themselves and thus increases the probability of detecting fish as anomalies.

A singular value decomposition (SVD) is computed on the collection of background vectors, $\mathbf{E} = \mathbf{U}\mathbf{S}\mathbf{V}^T$, where \mathbf{U} and \mathbf{V} are orthonormal matrices and \mathbf{S} is a diagonal matrix of singular values. The column vectors of \mathbf{U} contain a basis which best represents \mathbf{E} , with the singular values describing the energy that each vector of \mathbf{U} represents [387]. (Note that since \mathbf{E} was mean-subtracted, the columns of \mathbf{U} are the non-singular principle components. This saves on computation time, as the SVD is computed on the \mathbf{E} matrix, size $N_{p,r} \times N_e$, while the PCA is computed on the full covariance matrix of size $N_{p,r} \times N_{p,r}$ [106].) Only the first N_e non-zero singular values and associated vectors are required at most. The first few background eigenvectors associated with the singular values of highest energy are shown in Figure 2-24 for an example video captured aboard the *Ocean Foam*, reshaped back to the original size of $M_r \times N_r \times 3$.

The background estimate is formed for each new frame by applying the same median filtering and resampling, then decomposing the frame into the \mathbf{U} basis. Typically, only the

first $n_{bkg} = 4$ to 10 eigenvectors of \mathbf{U} are required to represent the changing background elements without describing elements associated with fish. Note that this again limits the number of eigenvectors from the SVD that need to be computed.

Predicting which frames are more likely to contain an interesting object limits the number of frames that must be passed through later, more computationally demanding processing steps. A linear discriminant analysis (LDA) provides a fast estimate of the probability that a frame may be interesting or blank by finding a single projection direction that maximally separates two labeled data classes. The projection direction vector, \vec{w} , is calculated as

$$\vec{w} = (\mathbf{\Sigma}_I + \mathbf{\Sigma}_B)^{-1} (\vec{\mu}_I - \vec{\mu}_B), \quad (2.7)$$

where $\vec{\mu}_I$ and $\vec{\mu}_B$ are the mean of the *interesting* frames and *blank* (non-interesting) frames, and $\mathbf{\Sigma}_I$ and $\mathbf{\Sigma}_B$ are the associated covariance matrices for each class [106],[24]. Taking the dot product of \vec{w} with a resampled frame gives a scalar value, c , for the data's position along the \vec{w} direction. Computing the distributions of c for the interesting and blank frames gives a way of estimating the probability that the c value for a new frame will have visual information worth processing.

The LDA makes a number of assumptions. One is that the data can be represented by a single multivariate Gaussian, so that the mean and covariance are adequate to represent the distributions for each class. Deviations from Gaussianity reduce the separability. In this instance, Gaussianity of the pixel values of each class is assumed to be a reasonable estimate. A second assumption is that the number of samples for each class is equal so that the uncertainty in the true covariance, as estimated by the sample covariance, is the same for both classes. A larger number of non-interesting samples biases the separation direction \vec{w} towards interesting frames. The \vec{w} direction can be corrected by including a $n_I^{1/2}$ and $n_B^{1/2}$ factor for each covariance matrix in 2.7, where n is the number of samples of each class type [106],[24]. For the purpose of predicting which frames may be interesting and avoiding false negatives, the bias is preferable.

One issue that arises during the LDA calculation is that the $\mathbf{\Sigma}_I$ and $\mathbf{\Sigma}_B$ matrices are rank-deficient. Each is $N_{p,r} \times N_{p,r}$ in size, but are constructed from n_I and n_B samples and thus have a maximum rank of n_I or n_B . In general, $N_{p,r}$ is an order of magnitude larger

than either n_I or n_B . The solution is to reduce the dimensionality of the data samples so that the covariance matrices are full rank [24],[189]. Since the SVD already contains a good representation of the expected images, the first $\min\{n_I, n_B\}$ vectors of \mathbf{U} are used to project the frames into a lower-dimensional subspace of the eigenspace. The covariance matrices in the \mathbf{U} eigensubspace are then full rank, assuming the data vectors are linearly independent (the \mathbf{U} matrix is orthonormal, so the transformation itself is linearly independent). At this point in the processing chain, the number of eigenvectors from \mathbf{U} which must be retained is $\max\{n_{bkg}, \min\{n_I, n_B\}\}$, where n_{bkg} is the number of eigenvectors from \mathbf{U} used to estimate the background. The remainder of \mathbf{U} can be discarded to save memory.

Labels for the frame data are estimated from the statistics of the frames. The Mahalanobis distance [80],[236],

$$D_M = \left[(\vec{p} - \vec{\mu})^T \mathbf{\Sigma}^{-1} (\vec{p} - \vec{\mu}) \right]^{1/2}, \quad (2.8)$$

is calculated for each RGB pixel sample \vec{p} , where $\vec{\mu}$ is the mean color at that pixel location and $\mathbf{\Sigma}$ is the color covariance matrix. If $\mathbf{\Sigma}$ is diagonal, the Mahalanobis distance reduces to a variance-normalized Euclidean distance. The scalar distance indicates the probability that the pixel belongs to the distribution described by $(\vec{\mu}, \mathbf{\Sigma})$; small distances indicate high probability of membership, while large distances indicate that \vec{p} is improbable. Frames with enough pixels above a pre-selected D_M threshold are considered to be *interesting*, while those with low D_M values are labeled as *blank*. This approach avoids the necessity of a human observer to provide the labels themselves.

Automatic labeling using the Mahalanobis distance provides a decision for *interesting* versus *blank* based on a fixed threshold, making a similar decision as from the LDA for whether further processing should be applied. The difference is that the LDA provides a probability for which hidden class which the frame belongs to based on its c value. This allows a Bayesian network to be constructed [106], assuming that the D_M labels are adequate for estimating the transition probabilities between the hidden classes. The solution of the Bayesian network then provides a better estimate of whether an interesting object exists in the frame by combining information from the neighboring frames and observations.

Finding specific regions in the individual frames marked for processing which are anomalous from the background was initially done using the Mahalanobis distance. Thresholding, morphological operations, and region processing were applied to determine candidate object patches. The main drawback was that fish which had a similar color to the background had small D_M values and poor edge gradients, so that the object detection would become disjointed. Improvements in the object segmentation were attempted using graph cuts [39],[238] and grow-cuts [401] on the patches, with region seeds estimated using both the Mahalanobis distance and object color statistics estimated using the candidate patches. Unfortunately, both methods suffered from significant overflow in the regions of color similarity and provided poor object segmentation. They also showed sensitivity to the region seeds and method parameters, and are not recommended without additional priors to constrain the cut regions—which require, in many cases, significant user intervention to generate [125],[239].

A statistical matching using the patches with high D_M proved to have better results in joining together similar regions. Candidate object patches which were near each were extracted from the image. The Hotelling’s T^2 statistic [163],[393], the multivariate generalization of Student’s t , was calculated on the color distributions between the two patches as

$$T^2 = \frac{n_1 n_2}{n_1 + n_2} (\vec{\mu}_1 - \vec{\mu}_2)^T \mathbf{S}^{-1} (\vec{\mu}_1 - \vec{\mu}_2), \quad (2.9)$$

where n_1 and n_2 are the number of samples in each patch, $\vec{\mu}_1$ and $\vec{\mu}_2$ are the mean color values. Matrix \mathbf{S} is the pooled covariance matrix composed from the two patch covariance matrices $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$ as

$$\mathbf{S} = \frac{(n_1 - 1) \mathbf{\Sigma}_1 + (n_2 - 1) \mathbf{\Sigma}_2}{n_1 + n_2 - 2}. \quad (2.10)$$

The T^2 statistic follows an F distribution with appropriate scaling for the number of samples and dimensions, which gives the probability that the two patches were sampled from the distribution of colors [393]. For those patch pairs with a low enough F -test and thus a high probability of being drawn from the same distribution, the convex hull around the patches was used to find the area connecting the two patches which was not originally detected. The entire set of pixels within the convex hull is then fit to a series of Gaussian mixture models



Figure 2-25: Three examples of fish which change appearance significantly between frames, making traditional tracking especially poorly suited for this application. Each view of a fish is taken from consecutive video frames at the recorded resolution.

(GMM) with one to three components using the Rissanen value (minimum description length) to determine the best number of mixtures [37],[208]. If one mixture is predicted, the three regions are grouped together into a single detection; if two or more mixtures are more likely, the candidate patches remain separate detections and the connecting region is ignored. The GMM is slower than the T^2 statistic to compute, but can combine all three regions with better reliability and noise tolerance.

2.3.2 Linking detections

Motion tracking traditionally relies on small and predictable changes in appearance between steps in time [149],[368]. Zhong et al. note that

“Many object tracking applications share the following properties: 1) the inter-frame motion of the object is small so that the object in the next frame is in the neighborhood of the object in the current frame, 2) the same point on the object has a consistent color/gray scale in all the frames of the sequence, 3) the boundary of the moving object has a relatively large motion field, and 4) the boundary of the object has a large image gradient value.” [443]

Low temporal sampling relative to the motion means that small deformations are unlikely (point 1) and that the color is likely to change as the fish rotates (point 2) (Figure 2-25). Color similarity to the background makes points 3 and 4 unreliable, as does the amount of motion from other objects in the region of interest.

Grouping detections from multiple frames using probabilistic models was used in deference to traditional tracking methods. The probabilistic approach allows for greater variation between each sample and does not rely on strong predictors for the motion field. First, the assumption is made that for two detections within a short time span of each other, the probability that they are from the same fish (or other object), p_s , is a function of the relative size, change in position, and color similarity. A second assumption is made that the functions for each descriptor are independent and follow an exponential decay, so that

$$p_s \simeq \frac{1}{Z} \exp(-\Delta f_{ab} T_{ab}^2) \exp(-\Delta f_{ab} l(\Delta \vec{x}_{ab})) \exp(-\Delta f_{ab} s(n_a, n_b)), \quad (2.11)$$

where Δf_{ab} is the number of frames between two detections a and b , T_{ab}^2 is the Hotelling statistic computed on the colors, $l(\Delta \vec{x}_{ab})$ is a function on the relative change in position between two detections, and $s(n_a, n_b)$ is a function which describes the relative size; the $1/Z$ term is for normalization. The exponential decay assumption is made because $\Delta f_{ab} T_{ab}^2$ follows nearly an exponential for real data. The other two terms also happen to have a nearly exponential decay for intuitive comparison functions: if $s(\cdot)$ is defined as

$$s(n_a, n_b) = \beta \left| \frac{n_a - n_b}{(n_a n_b)^{1/2}} \right|, \quad (2.12)$$

so that it measures a relative change in the size, $\Delta f_{ab} s(\cdot)$ also follows an exponential. Here, n_a and n_b are the number of pixels in the detections, and β is a scalar used to adjust the importance of this term. The distance similarity function, $l(\cdot)$, is set to be

$$l(\Delta \vec{x}_{ab}) = \alpha [\max\{\|\Delta \vec{x}_{ab}\|_2 - \Delta x_{allow}, 0\}]^{1/2}, \quad (2.13)$$

where Δx_{allow} is a threshold before a distance penalty is applied and α is an importance factor. The distribution of $\Delta f_{ab} l(\cdot)$ is rougher, but approximately exponential.

The p_s function is used in a graphical model to determine which detections belong to the same object by linking detections across frames. The problem is made more tractable by setting a fixed limit on the number of frames, Δf_{\max} , that can elapse without an object being detected before it is removed from contention. Periods of the video with no detections for

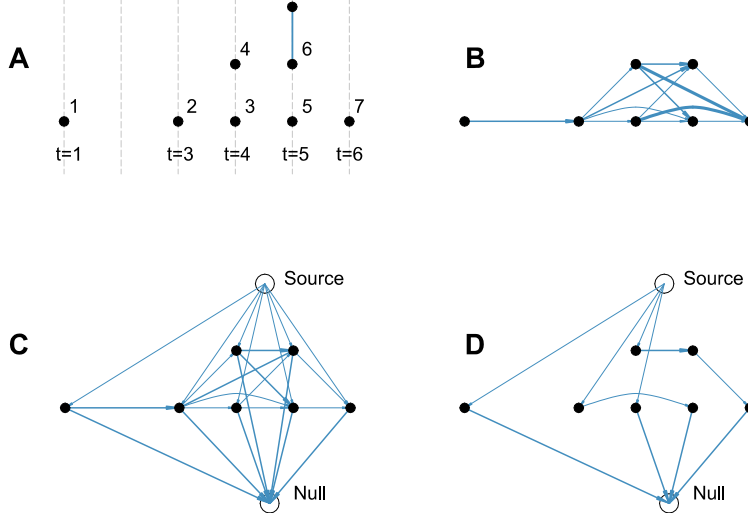


Figure 2-26: Example graphical model for linking together detections across frames. Detections are depicted as nodes, and arrows indicate that a later detection might have been generated from the object as the earlier detection. Line thickness is proportional to the dissimilarity between detections. Source and null super-nodes allow for a new object to appear into the view or an object to be removed.

at least Δf_{\max} are thus considered to be unrelated, forming separate graphical models. An example model is depicted in Figure 2-26, with nodes (black circles) representing separate detections. Panel A is a set of detections at each point in time, with two detections joined by the Hotelling/GMM method discussed in the previous section (detection 6). All possible links between detections at each frame are shown in panel B, where $\Delta f_{\max} = 2$ in this example and the line thickness indicates the dissimilarity. Two super-nodes are added to the graphical model, the source and sink of panel C, to allow objects to enter and leave the scene. Links to these super-nodes extend to every regular node.

The probability that an object enters the scene, produces the detections, then exits the scene is found by computing the most probable path over the graphical model. The probability, a product over the links, can be linearized by taking a log transform. The probability of two detections being generated by the same object, Equation 2.11, becomes a linear sum,

$$w_i = -\ln(Z) \Delta f_{ab} \cdot I_{\Delta f}(\Delta f_{ab} \leq \Delta f_{\max}) (T_{ab}^2 + \alpha l(\Delta \vec{x}_{ab}) + \beta s(n_a, n_b)), \quad (2.14)$$

where $I_{\Delta f}(\cdot)$ serves as an indicator function for whether the frame difference is less than the allowed maximum. The value w_i is the negative log-likelihood of p_s for link i , so that a low w_i indicates a strong similarity between $a \xrightarrow{i} b$. The weight for the links from the super-nodes is similarly the negative log-likelihood of the transition probability for an object to enter or exit the scene. In practice, the super-node weights are set to be higher than most of the node links but are set by some amount of trial and error since the normalization factor, Z , is not known *a priori* – and may change between different videos.

The best temporal linkage is found by solving a binary minimization problem over the graphical model,

$$\begin{aligned}
\min C &= \sum_i w_i e_i \\
s.t. \quad \sum_{i \in \{IN\}_j} e_i &= 1 \quad \forall j \\
\sum_{i \in \{OUT\}_j} e_i &= 1 \quad \forall j \\
e_i &\in [0, 1],
\end{aligned} \tag{2.15}$$

where e_i are binary decision variables, one per edge, for whether that edge was used as a link, $\{IN\}_j$ is the set of edges leading into node j , and similarly $\{OUT\}_j$ is the set of edges going out of node j . The setup forces each node to have exactly one incoming and one outgoing edge link, either between frames or to the source/null super-nodes. A solution of this type is depicted in panel D of Figure 2-26. It is worth noting that Equation 2.15 is a minimum/maximum flow problem [32], which is a well-studied problem and has a number of solution techniques [144],[76],[65],[32],[195].

The graphical model does an excellent job linking together similar detections between frames if the detection was reasonable. There are some issues when colors are estimated poorly or when objects overlap (such as a gloved hand grabbing a fish) that are difficult to overcome without using mixture models in the similarity functions. Examples of linked detections are shown in Figure 2-27.

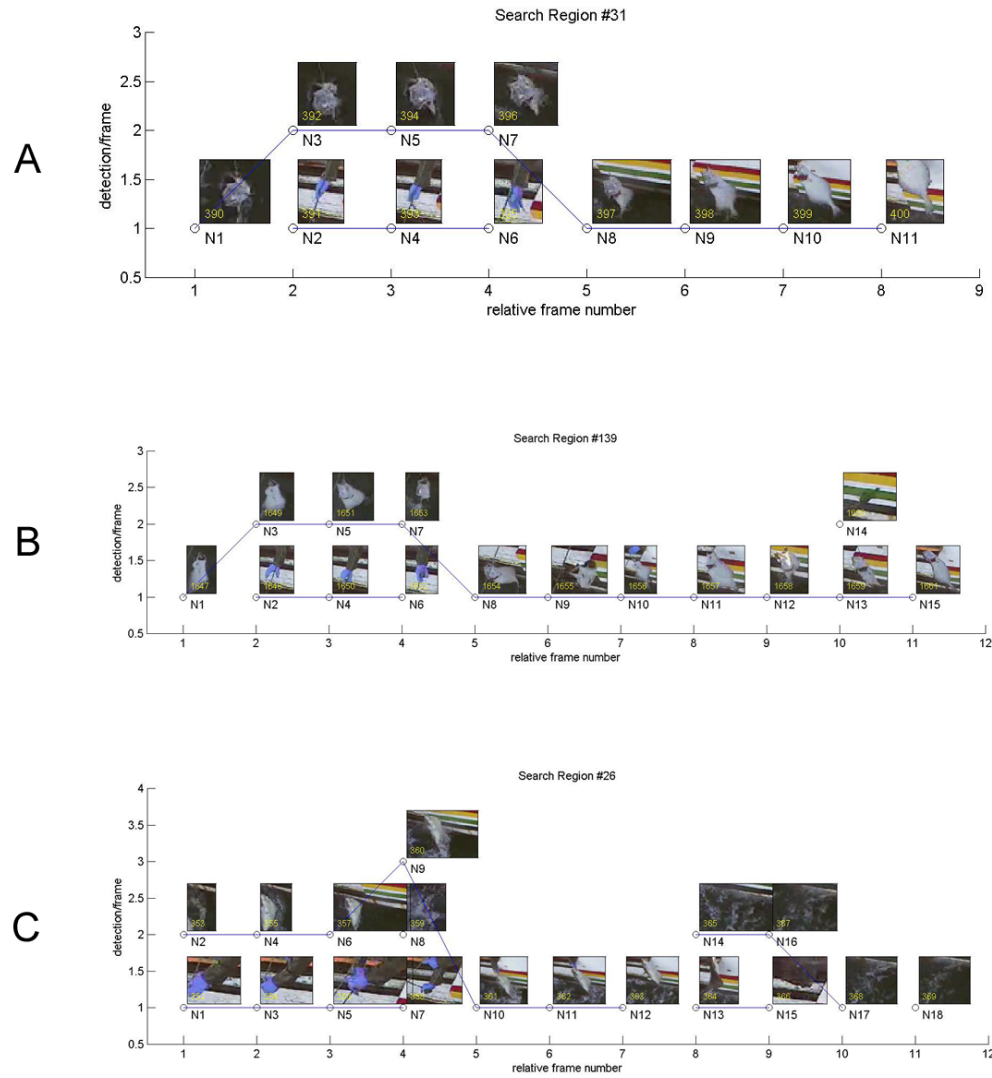


Figure 2-27: Examples of linked detections in three distinct periods of video. Areas around detections are shown above the nodes, and blue lines connecting nodes indicate the most probable links for the given settings. Examples A and B have correct links between the detections. Example C shows a case where the color and size change enough between frames 7 and 8 that the detections are not linked as a single object.

2.3.3 Improvement suggestions

The two greatest challenges of working with the longline video are the similarity between background and objects, and the low temporal resolution. Modifying the background so that it contains colors or textures that would not naturally occur in the fish species of interest would make extracting the regions significantly more reliable. Good extractions in a few videos would also allow a system to automatically learn shape and color priors, which could then be used for improved segmentation in videos that did not have modified backgrounds [125],[239],[66],[215] or to estimate distributions on motion for better probabilistic tracking [214]. A higher frame rate decreases the uncertainty in the estimation process, as the changes between frames are much smaller. A video recorded with lower spatial resolution (240×320 pixels, for example) but higher frame rate (10-15 frames per second) would take less disk storage space than the current size (480×640 at 5 fps) but offer more useful information for tracking and identification.

Setting correct thresholds for each step can be somewhat challenging, especially without priors to guide the self-discovery (such as predicting interesting frames from the LDA). One area where that could improve is through labeling the processed links. Matter et al. have demonstrated software for guided categorization that is able to learn the classification as the user sorts images, feeding back the known labels to create a better system on-the-fly [246]. Labels from a similar type of interactive classification/sorting system could be fed back to the graphical models to adjust α and β since the flow problem has fast solutions.

Chapter 3

Digital Holographic Instrument Development: Lab to Ocean

The microscopic scale provides the basis of the life in the ocean. The smaller plants and animals form the basis of the food chain for larger animals; charismatic examples include the baleen whales, adapted to filter plankton as their primary food source [78],[425], or many of the commercial fish stock such as cod and haddock that rely critically on plankton populations during their larval development [21],[74],[376],[304],[271]. The majority of the biomass is apportioned to the smaller scales [324], so that microscopic populations can have a significant impact on the greater oceans. For example, a significant disagreement in the oceanic nitrogen balances was partly solved through a detailed survey of bacterial trichodesmium colonies [87].

Studying marine life on the microscopic scale is particularly difficult. Traditional methods such as net sampling are able to sample a large volume, but at the loss of spatial resolution, destruction of fragile samples, and involved effort to re-sample and identify the collected species. Non-contact methods for *in situ* measurements revolve around optical and sonic methods. Sonar is unfortunately not able to identify species and has size limits based on the wavelength, so that while it can penetrate further in an aqueous environment, it does not offer the detail needed for species-level biological studies [172],[413],[421]. Optical approaches have the best chance at achieving species-specific measurements coupled with

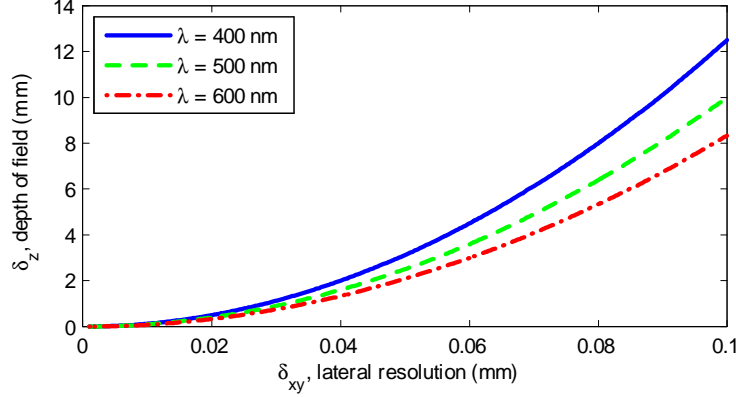


Figure 3-1: Tradeoff between lateral resolution and depth of field. The curve marks the maximum (approximate) limit that can be achieved based on $k = 1$ and assuming perfect imaging through the entire optical system. Real-world media limits standard optical systems to lie below these curves.

high precision regarding the location of each sample in the ocean.

Optical methods tend to be limited by size and resolution constraints. Lateral resolution, δ_{xy} , is given by

$$\delta_{xy} = \frac{k\lambda}{NA}, \quad (3.1)$$

where λ is the wavelength, k is a constant that depends on the imaging and resolution criterion, and NA is the numerical aperture of the imaging system. (The constant k ranges from 0.5 to 1.22 for various criteria and imaging modalities.) The corresponding depth of field (DOF), δ_z , is then

$$\delta_z \simeq \frac{\lambda}{2NA^2} = \frac{\delta_{xy}^2}{2k^2\lambda}. \quad (3.2)$$

Observing over a large volume with high lateral resolution thus requires a trade-off (Figure 3-1) [297]. Specialized optical systems can modify the DOF or resolution for fixed NA [102],[55],[444] but are still subject to a net loss [145]. The basic point remains that high-resolution imaging is incongruent with large imaging volumes.

Observing biology in an ocean environment adds physical separation between the measurement device and the sample. An instrument needs to not only journey to the sampling location but survive the trip, meaning that thick optical windows and housings are necessary. A few instruments rely on bringing samples closer to the imaging system by suctioning or

otherwise restricting movement of plankton [266],[311],[59],[179],[137]. Avoidance in these devices is likely, or the strategies for capture (i.e., a narrow slit that the plankton enter naturally [179]) are not scalable to zooplankton which have strong escape responses to fluid perturbations, such as copepods [47],[426],[168]. Systems such as the Video Plankton Recorder [88] are designed to reduce avoidance effects by using additional relay optics to transfer images from a volume further from the instrument housing to the recording camera. This has the result of increasing the size of the housings and relay lenses.

Digital holographic imaging is an ideal solution for *in situ* plankton measurements. It has a number of abilities and benefits:

- High spatial resolution on the order of microns to tens of microns is easily achievable, allowing for genus- or species-level identification and precise size measurements
- Holograms of large volumes can be recorded, on the order of hundreds of milliliters per exposure, important for sparse plankton distributions; the size of the optical volumes are well-characterized, important for concentration and size data
- The focal position can be computationally selected after the hologram has been captured
- Lens-less imaging is possible, significantly reducing the size of underwater housings
- Setups can be designed to be robust, compact, and flexible, with inexpensive and low-power components
- The hologram is recorded digitally, so that the data is immediately available for advanced digital image processing; similarly, the hologram can be stored to disk or archived and reconstructed at a later time with any number of algorithms and measurement goals (i.e., plankton identification, particle size distributions, characterizing water clarity, or estimating fluid flows)
- Orientation and 3D position of objects within the volume can be determined

This chapter discusses the development of a digital holographic imaging system suitable for *in situ* plankton studies. In particular, the system is designed for large sample volumes,

deep ocean deployments, and robust operation under harsh usage conditions. These three design goals already differentiate the new system from other holographic devices. The first half of the chapter briefly introduces digital holography, then discusses theoretical issues relevant to understanding the design and device performance. The second half of the chapter describes the development and practical implementation of the new holographic imaging system in a marine environment. Specific contributions of note include a Wigner analysis for understanding the properties and limits under spherical reference waves, modeling motion effects, the creation of an *in situ* prototype, and theory for understanding how oceanic particle distributions appear in holograms.

3.1 Introduction to digital holography

Digital holography (DH) is the title given to a number of related electronic techniques in holography [267],[205],[314],[315],[286]. In the past two decades it has come to mean that a hologram is recorded directly onto a digital sensor and the optical field is reconstructed computationally. Much of the previous work in holography since Gabor invented the technique in 1948 is applicable to fully digital systems with modifications for discrete sampling of the signal and a reduced space-bandwidth product (as compared to holographic films and plates). In addition to traditional approaches, techniques which take advantage of the digital representation for reconstruction and advanced filtering are now available.

A hologram is the coherent interference of an optical field diffracted (or reflected) from an object and a fixed beam used as a reference. A coherent laser beam is first shaped into a fixed reference wave, $r(x, y)$. Digital holography requires the reference to either be known via measurements or modelled. Reference waves with a planar phase,

$$r_p(x, y, z) = \exp(i[k_x x + k_y y + k_z z] + \phi), \quad (3.3)$$

or a spherical phase,

$$r_s(x, y, z_s) = \exp\left(\frac{i2\pi}{\lambda}\sqrt{x^2 + y^2 + z_s^2} + \phi\right), \quad (3.4)$$

are the two most common and easily formed references. (The “shape” is that of a surface of constant phase in the reference field.) The wave vector, $\vec{k} = [k_x, k_y, k_z] = 2\pi\hat{s}/\lambda$, gives the direction (pointing along \hat{s}) and wavelength, λ , of a propagating wave component. The depth, z_s , in the spherical phase reference is the distance from the source of the spherical wave to where it is measured. The sign of the phase is somewhat arbitrary [145], as is the offset of the phase, ϕ . (During hologram formation, the phase offset of the reference will be seen to cancel with the phase offset of the object’s reference component. Since the ϕ component cancels, it will be ignored for the sake of clarity.) These equations describe only the phase, leaving the amplitude and temporal components of the wave for later discussion.

The object wavefront is created when the reference interacts with an object in either reflection or transmission. A small object located at position z_a along the optical axis is modelled as a complex-valued transparency, $t(x, y) = A_t(x, y) \exp(i\phi_t(x, y))$, representing the absorption, A_t , and phase delay, ϕ_t , caused by an object. The field directly after the object is

$$a(x, y) = [1 - t(x, y)] r(x, y, z_a), \quad (3.5)$$

$$= r(x, y, z_a) - r(x, y, z_a) t(x, y) = r(x, y, z_a) + o(x, y; z_a), \quad (3.6)$$

written in the second form to explicitly show that the field is the sum of the reference, r , and an object wave, o . This field propagates some distance, z_p , to a second plane parallel to $a(x, y)$ (Figure 3-2) where the field is described by a convolution with a point-spread function (PSF), $h(x, y; z_p)$. The object wave is written as

$$o(x, y; z_p) = - \iint t(x, y) r(x, y, z_a) h(x - x', y - y'; z_p) dx' dy', \quad (3.7)$$

$$= [-t(x, y) r(x, y, z_a)] * h(x, y; z_p) \quad (3.8)$$

$$= o(x, y; z_a) * h(x, y; z_p), \quad (3.9)$$

where the $**$ symbol represents a two-dimensional convolution. The point-spread function is the equivalent of an impulse response in signal processing and describes how an infinitesimally small point source spreads out over space as it propagates. The convolution assumes

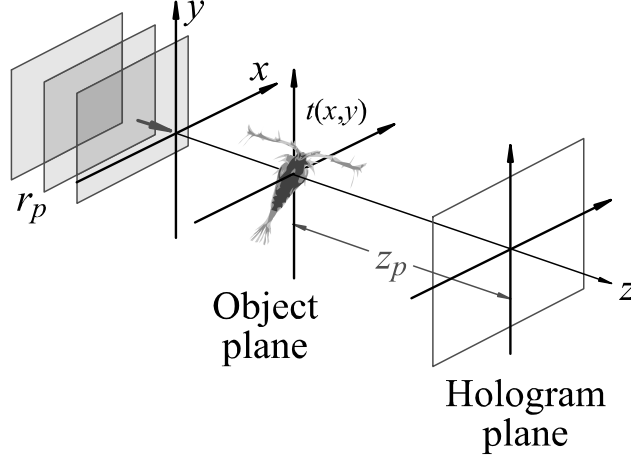


Figure 3-2: Coordinate system used in digital holography: x and y are lateral directions, and the signals propagate along the positive z direction.

that the PSF is spatially invariant and that the propagated object wave $o(x, y; z_p)$ remains linear [145], both common and easily satisfied assumptions. The latter requirement is only a concern for high energy fields [17]. The reference wave component can be written similarly by convolving $r(x, y, z_a)$ with $h(x, y; z_p)$, or more commonly, using knowledge of the reference itself; $r = r_p$ and $r = r_s$ both have analytical forms for any plane of interest.

The object and reference waves interfere at a detector plane to create an intensity pattern known as the hologram. The intensity is the time-averaged amplitude of the electric field and is given by

$$I(x, y, z) = |r(x, y, z) + o(x, y; z_p)|^2, \quad (3.10)$$

$$= |r|^2 + |o|^2 + r^*o + ro^* \quad (3.11)$$

$$= |r|^2 + |o|^2 + 2\text{Re}\{r^*o\}, \quad (3.12)$$

where $r + o$ gives the electric field, $|\cdot|^2$ is the squared amplitude of the complex field, and the conjugation forces I to be real-valued, encoding the phase information of the waves into the intensity amplitude [313],[145],[259]. The dependences of r and o on x , y , z , z_a , and z_p will be hidden for clarity until needed.

Equation 3.11 forms the basis of holography and deserves some discussion. The $|r|^2$

term is the amplitude of the reference wave and is often referred to as the “DC” term since it is nearly constant over (x, y) for a well-formed reference and does not change appreciably with the object. The $|o|^2$ term is the amplitude of the object wave, termed the “halo” due to the slight defocused glow it imparts to reconstructed fields. The amplitude of the diffraction is assumed to be weak compared to the reference wave so that $|r| \gg |o|$, realistic for small objects, so that $|o|^2 / |r|^2$ is negligibly small and o^*o is a small fraction of r^*o or ro^* . These last two terms, r^*o and ro^* , contain the object wave information modulated by the reference wave and are the image and “twin image” terms. The so-called twin is a secondary object image that appears to have been generated on the opposite side of the hologram plane thanks to the conjugation – though the working difference of which term is actually the image and which is the twin depends on preference for how the hologram is reconstructed.

Equation 3.11 hides a number of assumptions and details; for detailed discussion about the specifics of scalar diffraction theory, conditions for weak diffraction, and transparency functions, the reader is referred to the seminal texts of Goodman [145] and Born and Wolf [36]. The equation also assumes that the reference and object waves are coherent with each other, which depends on the coherence of the source, decoherence imparted by the propagation medium, and the optical setup. Holography with partial coherence is indeed possible, and is discussed in the literature [105],[104]; the assumption of high coherence is made here and is sufficient for the discussion in this chapter.

The “digital” part of digital holography refers to both recording the hologram on an electronic sensor and reconstructing the field computationally. A CMOS or CCD camera records the intensity directly, and the hologram can either be reconstructed immediately or stored on disk for later use. The key differences between camera types and models include a number of factors which are discussed in later sections of this chapter.

Reconstruction is performed computationally by deconvolving the hologram. Two simplifications will make the result of the deconvolution more practical. The first simplification is that the $|o|^2$ term in Equation 3.11 is negligibly small and can be ignored. The second is that the $|r|^2$ term can either be subtracted by assuming a known $|r|^2$, measuring the mean $|r|^2$ over a number of holograms, or by filtering the low spatial-frequency components. Al-

ternately, since the $|r|^2$ term acts as an offset in both the hologram and the reconstructed image, it does not negatively affect the actual reconstruction process. Assuming that the $|r|^2$ term is removed, however, clarifies the process. The intensity pattern after removing $|r|^2$ and $|o|^2$ is then

$$\tilde{I} = r^* o + r o^*. \quad (3.13)$$

Multiplying by the reference wave (or its conjugate; the same steps are applicable) and dividing by $|r|^2$ gives

$$\frac{\tilde{I}r}{|r|^2} = o + \frac{r^2 o^*}{|r|^2}, \quad (3.14)$$

the object wave at the detector plane, o , and a scaled term from the twin image. If the reference wave r is a plane wave aligned with the detector, $r^2/|r|^2 = 1$ and the hologram represents $o(x, y) + o^*(x, y)$. A plane wave at an angle to the detector results in a modulation of the o^* term by $\exp(-i2[k_x x + k_y y]) \exp(-i2k_z z)$; the $k_z z$ component becomes a constant phase offset. The effect with a spherical reference is similar, doubling the phase modulation on o^* .

Information about the object is contained in o and o^* , though it is obscured by propagation, Equation 3.9. Fortunately, removing the effect of propagation is possible because the PSF (the convolution kernel), h , can be derived from the physics of scalar diffraction theory as

$$h(x, y; z_p) = \frac{z_p}{i\lambda} \frac{\exp(ikd(x, y, z_p))}{d(x, y, z_p)^2}, \quad (3.15)$$

where $d(x, y, z_p) = [x^2 + y^2 + z_p^2]^{1/2}$ is the distance a ray¹ emitted from the point source travels to the point (x, y) on the plane a distance z_p away. For $z_p \gg x$ and $z_p \gg y$, the $d^2 \simeq z_p^2$ term is nearly constant in the denominator and the exponential can be approximated using the first three terms of its Taylor series expansion, yielding

$$h_f(x, y; z_p) = \frac{e^{ikz_p}}{i\lambda z_p} \exp\left(\frac{ik}{2z_p} [x^2 + y^2]\right), \quad (3.16)$$

the well-known Fresnel approximation. These two PSFs have direct Fourier transforms

¹The concept of a “ray” in the context of fields and waves is admittedly odd. In this case, a ray is likened to a small portion of the wave traveling in the same direction towards the same point.

given by

$$H(u, v; z_p) = \exp\left(\frac{i2\pi z_p}{\lambda} \left[1 - (\lambda u)^2 - (\lambda v)^2\right]^{1/2}\right), \quad (3.17)$$

$$H_f(u, v; z_p) = \exp(ikz_p) \exp(-i\pi\lambda z_p [u^2 + v^2]), \quad (3.18)$$

where u and v are the spatial-frequency variables corresponding to the x and y directions, respectively. (These functions are the equivalent of a transfer function in signal processing.) The equation for H is valid only when the argument in the square root is positive; outside this limit, the waves become non-propagating evanescent waves. The equation for H_f relies on the same assumptions as h_f , which in the context of the Fourier transform means that u and v correspond to plane wave components travelling at shallow angles relative to the optical axis. Goodman includes more detail on the derivation of these kernels and the necessary assumptions [145]. The numerical error in using H_f over the more exact H becomes extremely small when z_p is even a few millimeters, and has been examined by a number of earlier researchers (see, e.g., [217]).

The kernels of Equations 3.17 and 3.18 have a number of useful properties. One is that the product of j kernels can always be combined into a single kernel,

$$\prod_j H(u, v; z_j) = H\left(u, v; \sum_j z_j\right), \quad (3.19)$$

so that j propagations are equivalent to a single propagation over a distance $\sum z_j$. This immediately gives the inverse kernel, H^{-1} , as $H^{-1}(z_p) = H(-z_p)$, so that $HH^{-1} = \exp(0) = 1$; equivalently, $H^{-1}(z_p) = H^*(z_p)$. (The explicit dependence of H on u and v is hidden for the sake of clarity here and in following sections unless necessary for the exposition; rather, the propagation distance, z_p , is the variable of interest.) The physical interpretation is that the inverse kernel is a propagation in the negative (backward) z_p direction. In addition to sums, it will also be useful to note that the magnitude of H is constant over u and v , and that H is symmetric. The same properties hold for the Fresnel approximated kernel.

Deconvolution to reconstruct an optical field is done by multiplication in the spatial frequency domain with the inverse kernel. Taking the Fourier transform of the intensity,

3.14, gives

$$\hat{I}(u, v; z_p) = \mathcal{F} \left\{ \frac{\tilde{I}r}{|r|^2} \right\} = O(u, v; z_p) + \mathcal{F} \left\{ \frac{r^2 O^*}{|r|^2} \right\}, \quad (3.20)$$

$$= A(u, v) H(u, v; z_p) + \mathcal{F} \left\{ \frac{r^2}{|r|^2} \right\} ** A^*(-u, -v; z_p) H^*(-u, -v; z_p) \quad (3.21)$$

where $O(u, v; z_p)$ is the Fourier transform of $o(u, v; z_p)$ and $A(u, v)$ is the Fourier transform of $-t(x, y) r(x, y, z_a)$. Multiplying by the inverse kernel,

$$\begin{aligned} \hat{I}(u, v; z_p) H^{-1}(u, v; z_p) &= A(u, v) + \mathcal{F} \left\{ \frac{r^2}{|r|^2} \right\} ** A^*(-u, -v; z_p) \\ &\quad \times H^*(-u, -v; z_p) H^{-1}(u, v; z_p), \end{aligned} \quad (3.22)$$

$$= A(u, v) + \mathcal{F} \left\{ \frac{r^2}{|r|^2} \right\} ** A^*(-u, -v; z_p) H(u, v; -2z_p) \quad (3.23)$$

since $H(z_p) H^{-1}(z_p) = 1$ and $H^*(z_p) H^{-1}(z_p) = H(-z_p) H(-z_p) = H(-2z_p)$. The first term is the spectrum of the field just after the object and the second term is the twin image. The twin is the conjugate of the object, back-propagated by $2z_p$, and results in a severely out-of-focus object image if z_p is large compared to the object size. (The exact conditions are better quantified using Wigner relations developed later in this chapter, Section 3.2.) Since H has a constant magnitude, ill-posed division by small numbers is not a concern and Tikonov regularization [31] is not necessary². Taking the inverse Fourier transform of $\hat{I}(u, v; z_p) H^{-1}(u, v; z_p)$ gives the reconstructed optical field at z_p . Figure 3-3 shows an example reconstruction of a copepod, displaying the magnitude of the field. The twin image is visible as ringing around the object.

The second term can be simplified significantly by considering three common reference waves. If r is a plane wave with normal incidence to the detector, $r/|r| = 1$ and $\mathcal{F}\{r^2\}$ is a delta function centered at the (u, v) origin, leaving A^* unchanged. If r is a plane wave inclined to the detector, then $r^2/|r|^2 = \exp(i2[k_x x + k_y y])$ and the resulting delta function shifts A^* away from the origin. The third common reference is to set $r = r_s$, a spherical wave,

²Tikonov regularization and Wiener filters may be helpful in the case of strong noise with good priors on the noise power or the noise spectral distribution. Spectral windows, such as the power filtering discussed later in Section 3.3.4 in conjunction with removing the effects of Bayer color filters, is a version of Tikonov with priors based on the assumed distribution of replicated spectra.

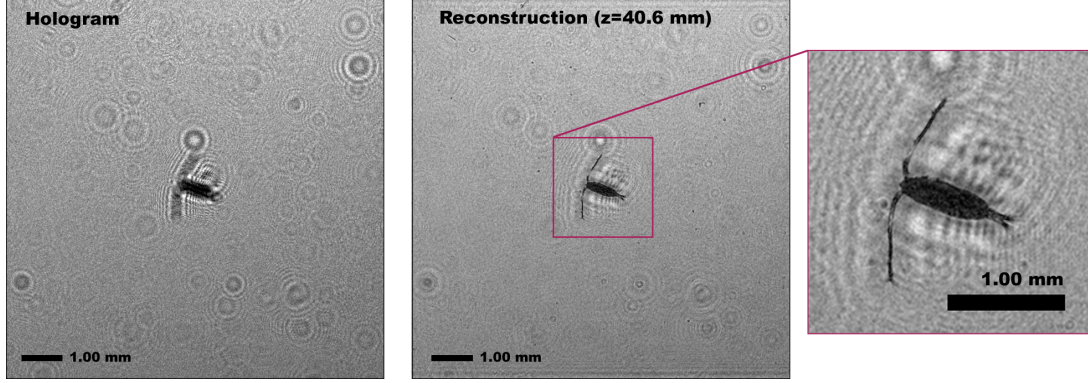


Figure 3-3: Example reconstruction of a digital hologram. Left: a 1024×1024 sample of the interference pattern recorded on the digital detector. Center: the magnitude of the optical field, reconstructed at a depth of $z_p = 40.6$ mm from the detector. Right: detail of the in-focus field at the object depth. Ringing around the object is due to the twin image. The reference has not been subtracted in this hologram and provides a bright background but otherwise does not affect the image.

which applies a spherical phase distortion to $o^*(x, y)$ in the spatial domain. The additional phase is the equivalent of passing o^* through a thin lens before back-propagating. Later in Section 3.2, a Wigner domain analysis show that in the case of spherical references, reconstructing $r^*o + ro^*$ is the better option to retrieve the transparency function $t(x, y)$.

Reconstructing optical fields through the Fourier-domain back-propagation approach includes a number of benefits. Both H and H_f have analytical Fourier transforms and can be calculated on-the-fly faster with modern hardware than loading pre-calculated kernels from a disk. (The difference between the two is that H is more exact than H_f for small z_p [217],[145], and the square root in H can be computationally expensive, around 2-5 times slower than a multiplication [344]. However, if multiple reconstructions are done, the square root term can be pre-calculated and stored to memory if the marginal savings are critical.) Additional Fourier-domain filters can be chained into the processing during the reconstruction. Multiple reconstructions can be done at an accelerated rate by noting that the initial Fourier transform of the hologram needs to be computed only once, then stored to memory for subsequent use with each new propagation distance. Reconstructions on planes angled with respect to the hologram plane are possible with a remapping of the spatial frequency data [245].

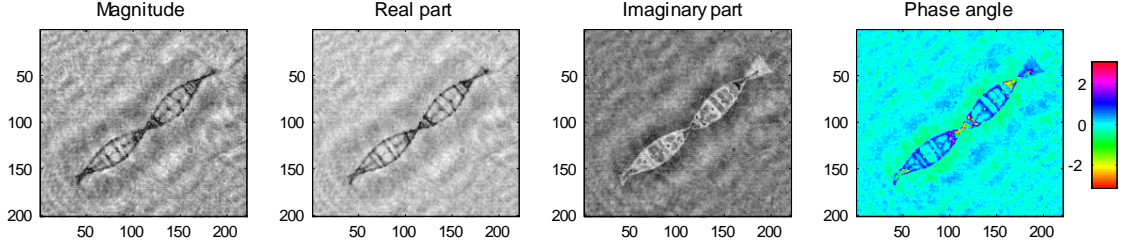


Figure 3-4: Components of the complex-valued reconstructed optical field for a mixed phase-amplitude object.

Other methods for reconstructing the field are possible and have uses for certain applications. These include compressive sensing approaches [40],[151], direct inversion [187],[349], wavelet decompositions [221],[222], or a Fresnel transform [115],[409]. The Fourier transform method is selected here for its balance between computational speed, memory usage, operational transparency, and the ability to chain multiple filters in a single operation. A large amount of prior work has shown that the method is more than sufficient for reconstructing images of plankton [217],[254],[95],[281],[366],[367].

The reconstructed field is complex valued, giving immediate access to more information than traditional holography (Figure 3-4). The real and imaginary components are useful as diagnostics, as components of focus metrics [89],[274], as representations of the objects, or for computing the quantitative phase of imaged objects [75]. The squared magnitude gives the intensity which would be observed using optical reconstruction methods.

The propagation and reconstruction depths in this section assume an inverse optical path length, as opposed to a physical length. The inverse optical path length (IOPL) is

$$IOPL = \int_0^L \frac{1}{n(x)} dl, \quad (3.24)$$

where n is the real part of the index of refraction and the integral is over the physical path length. For $n > 1$, the IOPL represents an optically foreshortened distance [136],[161],[113]. This is especially important for oceanic holography, where the waves pass through water with an index of refraction $n = 1.33 - 1.35$ [291],[256] and at least one optical window with n ranging from $\sim 1.4 - 1.8$ [155]. Care should be taken to use the IOPL to specify the recording and reconstruction parameters, then back-correct using the geometry and indices

of the particular setup if the physical distance is needed. Of particular note, the distances given throughout the remainder of this chapter and the next are the inverse optical path lengths unless stated otherwise.

The actual index variation within aqueous environments tends to be small enough over the scale of a device that distortions of objects are not noticeable. (Conversely, the biologicals imaged have enough intrinsic shape variations that distortions were usually not recognizable.) Exceptions include mixing regions, where salinity, density, and temperature may be all be fluctuating within the image volume. If distortions are a concern, Tippie et al. suggest methods for estimating and correcting certain phase distortions given some assumptions about the object image [381],[378].

The introduction here is meant to give the reader a working understanding of digital holography and is not intended to be a complete survey of the field. A number of excellent books and references on the subject are available for those interested in the finer points. In particular, Goodman [145] and Born and Wolf [36] cover the foundations of scalar diffraction theory; Schnars and Jüptner [313],[316], Kreis [202], Poon [287], Milgram and Li [254],[217], and Domínguez-Caballero [95] provide more detail regarding the general application of digital holography.

3.1.1 Modeling motion during exposure

Understanding the effects of motion on hologram formation describes what to expect from vibrations and object movement during the exposure, of particular interest for *in situ* sampling of active biological objects. The observed intensity at the detector plane is proportional to the time average of the squared electric field magnitude [36],

$$I \propto \langle E^2 \rangle = \langle (r + o)(r + o)^* \rangle, \quad (3.25)$$

which gives rise to Equation 3.11 if r and o are considered to be constant and the expectation value is thus constant. (The term “constant” in this context means that the spatial component of the waves is not changing over the integration time.) The period of the time averaging is taken to be longer than the electric field oscillation period. Consider a simple

case where $g = \cos(\omega t)$; the time average over a time period of length T is then

$$\langle g^2 \rangle = \frac{1}{T} \int_0^T \cos^2(\omega t) dt = \frac{1}{2} \frac{\cos(\omega T) \sin(\omega T) + \omega T}{\omega T}, \quad (3.26)$$

and the relative error of stopping the integration at $T = \tau_s$ is $\cos(\omega \tau) \sin(\omega \tau) / \omega \tau$. The numerator has a maximum possible value of 0.5, so that the maximum error is $0.5 / \omega \tau$. For visible wavelengths, $\tau_s \gtrsim 1$ ps is long enough to reduce the error to a tiny fraction of a percent, so that r and o need only be spatially constant over τ_s . The total energy measured at the detector is the sum of the intensities over many short moments in time – in a sense, the sum of picosecond moments³. A hologram is thus the incoherent sum of these short-time intensities [94].

Consider now the specific case of an object which translates laterally over the exposure time but does not change its transparency function, such as objects carried by water flowing quickly through a holographic volume. The field just after the object during period m is

$$a_m(x, y) = [1 - t(x - x_m, y - y_m)] r(x, y, z_a), \quad (3.27)$$

$$= [1 - t(x, y) * \delta(x - x_m, y - y_m)] r(x, y, z_a), \quad (3.28)$$

where x_m and y_m are the spatial shifts of the object transparency. The δ function acts to shift the transparency in space over the motion path. The resulting object wave during period m is

$$o_m(x, y; z_p) = (-r(x, y, z_a) [t(x, y) * \delta(x - x_m, y - y_m)]) * h(x, y; z_p). \quad (3.29)$$

If an in-line, planar reference is assumed, $r(x, y, z_a) = |r_p|$ and the constant-valued reference can be removed from the convolution. (This simplification makes use of the fact that the $\exp(ikz + \phi)$ plane wave phase offset cancels during hologram formation, so the phase can

³A more conservative measure of “moment” is found by noting that light takes around 1.5 ns to travel a 0.5 m optical distance, the average distance between the laser and detector in this chapter. Thus, photons which are being absorbed by the detector at $t = 0$ cannot interact with photons impacting the detector at $t = 1.5$ ns, since those later photons had yet to be emitted from the laser. This time period is three orders longer than τ_s , but guarantees (quantum mechanics aside) that the intensities received during these periods can be summed incoherently. Experiments with nano-second pulses [225] further reinforce the concept.

be ignored. Alternately, ϕ can be chosen arbitrarily so that $\exp(ikz + \phi) = 1$.) Applying the associativity of convolutions, o_m becomes

$$o_m = -|r_p| t(x, y) ** h(x, y; z_p) ** \delta(x - x_m, y - y_m), \quad (3.30)$$

$$= o(x, y; z_p) ** \delta(x - x_m, y - y_m), \quad (3.31)$$

where $o(x, y; z_p)$ is the object wave that would result from a stationary object.

The observed hologram is the sum of holograms from each period,

$$\begin{aligned} I(z) &= \sum_m (r_p(z) + o_m(z)) (r_p(z) + o_m(z))^*, \\ &= m|r_p|^2 + \sum_m |o ** \delta_m|^2 + r_p^* \sum_m (o ** \delta_m) + r_p \sum_m (o^* ** \delta_m), \end{aligned} \quad (3.32)$$

$$= m|r_p|^2 + \sum_m |o ** \delta_m|^2 + (r_p^* o + r_p o^*) ** \sum_m \delta_m, \quad (3.33)$$

where $\delta_m = \delta(x - x_m, y - y_m)$ is the shifting function for period m . The sum of deltas, $\sum_m \delta_m$, is a collection of points along the motion path. As m becomes large, the sum approaches a smooth curve through (x, y) space with intensity proportional to the time spent at that location, the equivalent motion point-spread function. Noting that the convolution of the constant $|r_p|^2$ with any real-valued point-spread function remains proportional to $|r_p|^2$, the hologram with motion can be written as

$$I = \left(|r_p|^2 + |r_p| o + |r_p| o^* \right) ** \sum_m \delta_m + \sum_m |o ** \delta_m|^2. \quad (3.34)$$

An approximation, I_L , to I is found by convolving a stationary hologram with the lateral motion PSF, giving

$$I_L = \left(|r_p|^2 + r_p^* o + r_p o^* + |o|^2 \right) ** \sum_m \delta_m, \quad (3.35)$$

$$= \left(|r_p|^2 + r_p^* o + r_p o^* \right) ** \sum_m \delta_m + |o|^2 ** \sum_m \delta_m, \quad (3.36)$$

so that the only difference between the actual hologram and the motion-approximated hologram is in the halo term (Figure 3-5). The assumption that $|o|^2$ is negligible compared

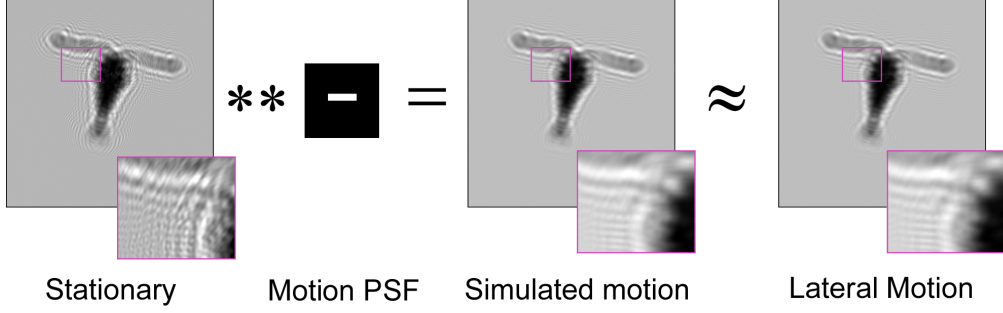


Figure 3-5: Simulating lateral motion of the object by blurring a stationary hologram. Left: stationary hologram. Center: I_L , the approximation found by convolving the stationary hologram with a sum of δ functions (the motion PSF). Right: the motion hologram I . The difference between I and I_L is on the order of the machine precision for this example.

to $|r_p|^2$ makes the error negligible. Image reconstruction works similarly, in that the image of a moving object is nearly equivalent to the image of the stationary object convolved with the motion PSF, aside from the halo term. Understanding how lateral motion affects holographic reconstructions is then a matter of understanding how the motion PSF affects a general image⁴.

The next case to consider is axial motion of the object, again assuming that $t(x, y)$ is not changing significantly over the exposure. The field after the hologram is

$$a_m = [1 - t(x, y)] r(x, y, z_{a,m}), \quad (3.37)$$

where $z_{a,m}$ is the axial position of the object at the m^{th} period. The field propagates a distance $z_{p,m}$ to the detector at time m , where $z_{p,m} = z_p + (z_{a,0} - z_{a,m})$ and z_p is the distance between the object and the detector at $m = 0$. The electric field at the detector is

$$E = r(x, y, z_{a,m}) * h(x, y; z_{p,m}) + [-t(x, y) r(x, y, z_{a,m})] * h(x, y; z_{p,m}),$$

which can be simplified if the reference is again an in-line plane wave, so that $r(x, y, z_{a,m}) =$

⁴One recent example comes from Reference [94], in which the authors use the blurring of a hologram to estimate particle velocities in the lateral direction. The “phenomenological relationship” of their Equation (3) actually arises from a short, linear motion blur, where their v/v_o relates to the blur length.

$|r_p| \exp(ikz_{a,m} + \phi)$. Substituting the Fresnel approximation for h gives

$$r_m = \frac{|r_p| \exp(ikz_{p,m})}{i\lambda z_{p,m}} \exp(ikz_{a,m} + \phi), \quad (3.38)$$

$$o_m = -\frac{|r_p| \exp(ikz_{p,m})}{i\lambda z_{p,m}} \exp(ikz_{a,m} + \phi) \left[t(x, y) ** \exp\left(\frac{ik}{2z_{p,m}} [x^2 + y^2]\right) \right] \quad (3.39)$$

for the reference and object wave components of E . The phase components will again cancel in each of the m different $r_m o_m^*$ and $r_m^* o_m$ terms of the hologram, despite the non-equal $z_{a,m}$, so that the phase can be ignored. Assuming that the motion is small so that $z_{p,m} \simeq z_p$, the hologram is then

$$I = m |r_p|^2 + \sum_m |o_m|^2 + \frac{|r_p|}{\lambda z_p^2} \sum_m [t(x, y) * h'_f(z_{p,m}) + t^*(x, y) * h_f^*(z_{p,m})], \quad (3.40)$$

with the prime on h'_f denoting that it has the same chirp component as h_f but without the constant phase offset. If a further assumption is made that $t(x, y)$ is real-valued (i.e., has only absorption), then I becomes

$$I = m |r_p|^2 + \sum_m |o_m|^2 + \frac{2|r_p|}{\lambda z_p^2} t(x, y) ** \sum_m \cos\left(\frac{k}{2z_{p,m}} [x^2 + y^2]\right). \quad (3.41)$$

The effect of the summed cosines is to average out the higher spatial frequencies. The Fresnel approximation was used here for the sake of making the phase cancellation readily apparent; the exact PSF gives essentially the same result, but with a square root in the cosine argument and marginally different intensity scaling.

The $h(z_{p,m})$ PSFs can also be written as a chained convolution,

$$h(z_{p,m}) = h(z_p) * h(z_{a,0} - z_{a,m}). \quad (3.42)$$

Using the same assumptions as before, the hologram can be factored as

$$I = m |r_p|^2 + \sum_m |o_m|^2 + \frac{|r_p|}{\lambda z_p^2} t(x, y) ** \sum_m [h'_f(z_p) * h'_f(z_{a,0} - z_{a,m}) + h_f^*(z_p) * h_f^*(z_{a,0} - z_{a,m})] \quad (3.43)$$

If the axial velocity is constant during the exposure, this last equation can be simplified by a change of variables: z_p is set to be the propagation distance at the middle of the exposure, and the object moves a total distance of z_r . The propagation PSF then becomes

$$h(z_{p,m}) = h(z_p) ** h\left(-z_r/2 + \frac{z_r}{M}m\right) \equiv h(z_p) ** h_{z_r,m}, \quad (3.44)$$

where M is the total number of time periods under consideration. The sum of propagation convolutions becomes

$$\sum_m h(z_{p,m}) + h^*(z_{p,m}) = \sum_m [h(z_p) ** h_{z_r,m} + h^*(z_p) ** h_{z_r,m}^*], \quad (3.45)$$

$$= [h(z_p) + h^*(z_p)] ** \sum_m h_{z_r,m}, \quad (3.46)$$

which comes from noting that $h_{z_r,m} = h_{z_r,M-m}^*$. Substituting back into the hologram formation equation,

$$I = m|r_p|^2 + \sum_m |o_m|^2 + \frac{2|r_p|}{\lambda z_p^2} t(x, y) ** \cos\left(\frac{k}{2z_p}[x^2 + y^2]\right) ** \sum_m h_{z_r,m}. \quad (3.47)$$

This form makes it apparent that motion in the axial direction is approximately the same as convolving the stationary hologram with $\sum_m h_{z_r,m}$.

The $h_{z_r,m}$ sum becomes an integral in the limit. The integral in the spatial domain, however, is not particularly illuminating. Instead, taking the Fourier transform and evaluating the integral gives

$$\begin{aligned} \int_{-z_r/2}^{z_r/2} \exp(i\pi\lambda z\rho^2) dz &= \frac{\exp(\frac{1}{2}i\pi\lambda z_r\rho^2) - \exp(\frac{-1}{2}i\pi\lambda z_r\rho^2)}{i\pi\lambda\rho^2}, \\ &= \frac{2\sin(\frac{1}{2}\pi\lambda z_r\rho^2)}{\pi\lambda\rho^2}, \\ &= z_r \operatorname{sinc}\left(\frac{1}{2}\lambda z_r\rho^2\right), \end{aligned} \quad (3.48)$$

where $\rho^2 = u^2 + v^2$ and (u, v) are the Fourier-domain spatial frequency variables. This last equation makes it clear that the axial motion reduces the high spatial frequencies of the

hologram.

Equation 3.48 also gives a simple limit for z_r so that no spatial frequencies are lost by requiring the zeros of the sinc function to be located beyond the detector's spatial frequency band. The first zero of the sinc occurs when $\frac{1}{2}\lambda z_r \rho^2 = 1$. Defining the detector's maximum frequency to be $\rho = \rho_{\text{det}}$, the motion range should be constrained to $z_r < 2/\lambda \rho_{\text{det}}^2$ to force the zero outside the detector bandwidth. This is twice the maximum depth of field for a Nyquist-limited sensor (see Section 3.2.4 and noting that the δ_{DOF} defined in that section is a unidirectional measurement) – though as will be seen later, the depth of field condition can be relaxed based on the object's spatial frequency content. What can be expected is that the fine-detail contrast will decrease slightly for objects with high frequency content as their total axial motion shift approaches the depth of field. Objects with low spatial frequency content should be able to translate further without visible degradation of the image.

The same types of arguments can be applied to motions of the detector and the reference wave source when using a planar reference. Lateral and axial motion of the source is particularly simple to handle, as it only modifies the phase offset of the wave. As noted during the discussion here, the phase offset cancels during in-line hologram formation⁵ and translational motion of the source should be impossible to measure. Motion of the detector is the same as for a moving object, a matter of changing physical reference frames. Lateral motion of the detector results in blurring along the motion path. Axial motion reduces high frequency detail, though it may not be noticeable if the displacement during the exposure is on the order of the depth of field – typically around 200 μm for modern detectors and visible wavelengths.

Tilting a planar reference beam slightly away from the normal makes little difference to the hologram formation. Matsushima et al. have derived relations for the fields on two arbitrarily aligned planes [244],[245] (see also Delen and Hooker, [91]). Of interest here is the relation between the field at a hologram plane normal to the optical axis and one that has been rotated by a small angle φ about the y axis. They shown that the spectrum in

⁵Phase-shifting holography necessarily relies on an independently controlled reference beam for this reason.

the rotated plane, $G(\hat{u}, \hat{v})$, is related to the spectrum in the un-rotated coordinate system $F(u, v)$ by

$$G(\hat{u}, \hat{v}) = F(\hat{u} \cos \varphi + \hat{w}(\hat{u}, \hat{v}) \sin \varphi, \hat{v}) \times \left| \cos \varphi - \frac{\hat{u} \sin \varphi}{\hat{w}(\hat{u}, \hat{v})} \right|, \quad (3.49)$$

where

$$\hat{w}(\hat{u}, \hat{v}) = [\lambda^{-2} - \hat{u}^2 - \hat{v}^2]^{1/2} \quad (3.50)$$

is the component of the wave vector in the \hat{z} direction of the tilted plane (the normal to the tilted plane). The second term in Equation 3.49 is the Jacobian and corrects the energy of the transformed waves; Matsushima notes that for small φ , the Jacobian is approximately constant. Holography typically has small u and v compared to w , and consequently $\hat{w} \sin \varphi$ is considerably larger than $\hat{u} \cos \varphi$. This results in a noticeable shift in the center of the spectrum from u_0 to \hat{u}_0 . Fortunately, a shift in the Fourier domain is equivalent to multiplying the spatial function by a complex exponential—which for optics is a plane wave. Following exactly the same arguments as earlier, this plane wave factor will cancel out during the hologram formation since both the r and o terms include this same phase multiplier. Thus, the G spectrum can be shifted back by \hat{u}_0 for comparison against F to observe how the spatial frequencies in the hologram formed by F would be mapped into the hologram formed by G .

Figure 3-6 plots the remapping function between \hat{u} in the tilted plane and u in the original plane. (As noted, the \hat{u}_0 shift has been included here for transparency in visualizing how reconstructed image frequencies are remapped.) The lines are calculated for a detector with 9 μm pixel pitch and a wavelength of 658 nm, and only the central line at $v = \hat{v} = 0$ is displayed; non-zero \hat{v} have nearly negligible influence since \hat{v}/\hat{w} is 0.036 or less. For even a relatively large $\varphi = 0.25$, the difference between u and \hat{u} is extremely small, around 2 mm^{-1} at the highest spatial frequencies. The mapping is nearly linear, varying by a small fraction of the spatial frequency. The end result is that for small φ , the hologram formed on a tilted detector is nearly identical to the original hologram. The reconstructed image should have a slight demagnification along the x direction due to the $\cos \varphi$ factor, but remains a faithful reproduction of the original object. The image can be corrected by resampling in the spectral domain [90] or in the spatial domain if φ is small and the non-linearity of \hat{w}

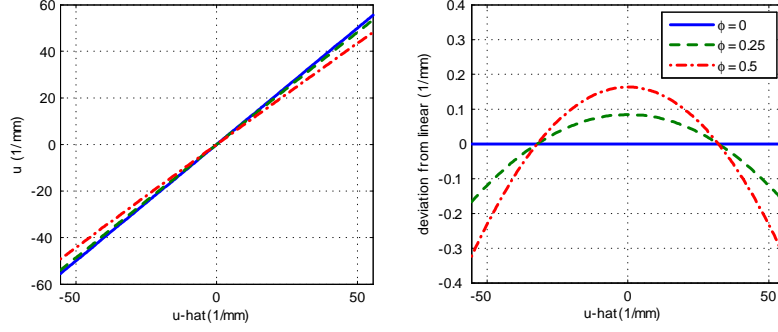


Figure 3-6: Coordinate transformation between \hat{u} and u . Left: the values of the spatial-frequency remapping for three angles; the lines vary by only a few mm^{-1} at most. Right: the non-linearity of the remapping function, measured as the deviation from a linear fit. The amount of deviation is extremely small over the range of spatial frequencies under consideration.

can be ignored.

Rotational motion of the detector leads to a blurring of the image in the rotation direction by remapping the spatial frequencies over a small range of \hat{u} . The highest spatial frequencies have the most shift during the remapping, around $u(1 - \cos \varphi)$ for small φ . Maintaining this frequency shift to be on the order of Δu , the sampling in the frequency space, means that $(1 - \cos \varphi) \sim 1/N$, where N is the number of samples. While this is a rough approximation and says little for the change in the actual image, it gives a useful rule-of-thumb limit: for $N = 1024$, φ is around 2.5° .

Spherical reference waves have many of the same results when the motion is small. In particular, the rotation of the detector results in the same spatial frequency remapping. Axial motion includes a slight change in a magnification factor M , where the magnification is given by Equation 3.63 and any losses due to blurring are highly object dependent. The spherical source has rotation invariance if the rotation is around the virtual (or real) point source.

The general conclusions for this section are two-fold. One is that a number of different types of motion can be approximated by convolving the stationary hologram with a kernel representing the motion PSF. Some types of motion, such as axial movement, may be negligible if the total change in position is small over the exposure time. General motion, including objects with changing transparency functions, can be modelled by incoherently

summing a number intensity holograms. The second conclusion is that in-line holography is especially robust to changes in the detector and source positions, both translational and rotational, and robust against axial motion within the imaging volume. Shorter exposure times make the hologram less susceptible by decreasing the distance that any one component or object can travel during that time period. Planar and spherical references have invariances which can be useful for robust designs.

3.1.2 Minimum number of bits

Holograms are particularly resilient to errors because the information about any one point is distributed across a larger spatial extent as the signal propagates. One consequence is that a reduced number of bits is required to store reasonable information about an object. For example, Domínguez-Caballero demonstrated that images can be reconstructed from the most limiting case of two binary levels, though with a significantly reduced clarity [95]. Figure 3-7 shows two examples of a simulated test object, a stylized “MIT”, for the case of a binary hologram and a two-bit (four-level) hologram. While the quality is lacking, there is a surprising amount of detail available.

The required bit count affects the engineering choices in a number of areas. The two most obvious are that a detector needs to have a high enough dynamic range to record the minimum number of bits in the first place, and that it has a high enough sensitivity to do so during the laser strobe period. The hologram signal also needs to be balanced against the background within the limits of the detector’s saturation level.

Estimating the minimum number of bits necessary for a reconstruction requires a metric for image quality and a ground truth. The ground truth is provided by simulated holograms computed with a sufficiently high number of bits. A number of metrics could be used for comparison of the image quality. The family of squared error metrics includes mean squared error, sum squared error, root mean squared error, and the normalized mean square error, and all describe different transforms on the statistical variance of the difference [116]. Squared error metrics are themselves a subset of ℓ_p norm metrics, of which the ℓ_1 and ℓ_∞ are used for mean error and maximum error, respectively; a fractional p allows interpolation between sum squared error ($p = 2$) and a sum of the absolute error ($p = 1$). A proposed

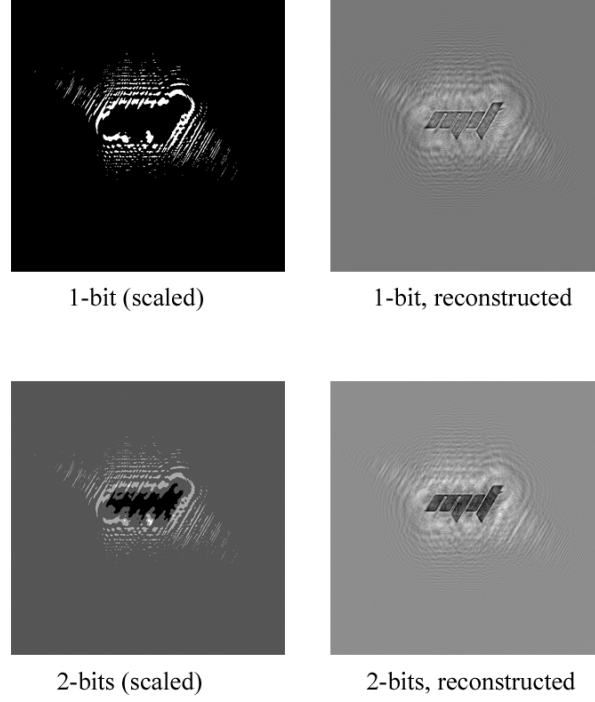


Figure 3-7: Example reduced bit-count holograms and the resulting reconstructed images.

metric using singular value decompositions measures the spectral differences between images [333]. Other recent methods focus on comparing the structure and various components which are preferred by the human visual system, but are not used here due to their lack of quantifiable relationship to the digital imaging and object recognition problem.

The normalized mean squared error (NMSE) was selected for its invariance properties and for its relation to the signal-to-noise ratio (SNR). It is defined by Fienup as [116]

$$NMSE \equiv E^2 = \frac{\sum_i (f_i - \alpha g_i)^2}{\sum_i f_i^2}, \quad (3.51)$$

where f is the ground truth image, g is the measured image, i is the discrete pixel index. Because g may have an unknown scale factor, α is used to compensate, and is chosen to be

$$\alpha = \frac{\sum_i f_i g_i}{\sum_i g_i^2}, \quad (3.52)$$

which minimizes the sum squared error and thus provides the best possible match between the expected signal and the measured data. The NMSE is related to the SNR since the

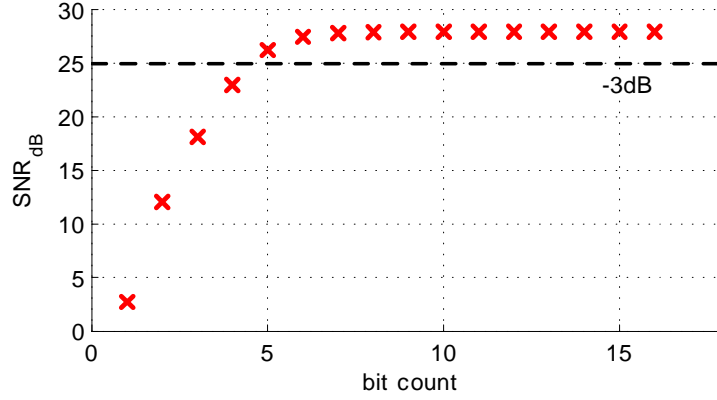


Figure 3-8: Approximate SNR of reconstructions from holograms with reduced bit counts

numerator in 3.51 essentially represents the noise power and the denominator the signal of interest, so that

$$SNR \approx \frac{1 - \langle NMSE \rangle}{\langle NMSE \rangle}, \quad (3.53)$$

valid for SNR away from unity and assuming that the noise is approximately uncorrelated with the signal.

Simulations were computed using an ensemble of 85 test objects to create digital holograms. Reconstructions using holograms with reduced bit counts were compared against the reconstructions from the original 32-bit digital holograms. The approximate SNR computed from the ensemble average NMSE is plotted in Figure 3-8, with the -3dB level denoted by the thick dotted line. The -3dB line, while somewhat arbitrary, corresponds well with human observations that the reconstructed image does not change significantly above about five bits. Moreover, signals with fewer bits still have a surprisingly high SNR. For engineering decisions, a minimum limit of around 4.5-5.5 bits is recommended – with the caution that this simulation does not include additional noise beyond discretization errors of the recorded hologram. Additional bits would be necessary to overcome significant detector noise or random noise due to the medium or laser source.

3.2 Wigner transform analysis for spherical reference waves

Holograms recorded with planar reference waves have been well-studied and understood due to their simplicity. On-axis planar references reduce to constant-valued scale factors, while off-axis planar references give a sinusoidal modulation to the object wave and thus act as carrier waves. These effects and their consequences are described in detail in the literature [145],[313].

The topic of spherical reference waves in holography has appeared at various times in the literature (see, e.g., [145],[313],[433],[137]), but has seen limited analysis as the math becomes more complicated—typically showing how to reconstruct an image and demonstrating the change in image size (and thus resolution). This section seeks to explain the limits of a holographic system which uses a spherical reference wave, and differs from previous analyses in its use of Wigner transforms to make the mathematics intuitive. In addition to explaining image size and resolution limits [227], the Wigner formulation can also be used to explain the depth of field and volumetric sampling [230] as functions of the object and detector bandwidths. The results and methods in this section are applicable to both spherical and planar references by taking the limit as the spherical source position goes to infinity.

The discussion here is more than academic. A good understanding of hologram formation and useful limits under spherical references enables their use in real-world applications, such as the *in situ* prototype of Section 3.3. It will also be seen that spherical references have a number of equivalences with planar references, useful for the laboratory holographer. Axial misalignment of collimators, lens aberrations, finite point-sources, and non-uniform beams (particularly, Gaussian) cause approximately spherical phase deviations, so that the ideal plane wave is actually a shallow spherical wave. Holograms can still be recorded and reconstructed robustly without a second thought from most experimenters despite these imperfect planar references. The derivations in this section will show why reconstruction with a planar reference does not affect the image and why most users of digital holography will never know, or need to know, that their reference is actually spherical.

3.2.1 Wigner distribution and optical formulation

The Wigner transform is a powerful analytic tool that describes the frequency content of a signal at a particular point in space, one of the Cohen class of space-frequency distributions [19],[20]. In this respect, the Wigner transform shares some similarities with short-time Fourier transforms [8],[340], S-transforms [360],[359], and wavelet analyses [362]; one difference is that the Wigner does not use any windowing. It also happens to have a number of particularly useful properties for optics [20]. This discussion will make use of the Wigner transform's ability to convert diffraction into geometric transforms. Understanding how an optical system affects a particular component of a signal is a matter of applying the geometric transform.

The Wigner transform is defined for a 2D signal $f(x, y)$ as

$$W(x, y, v, \omega) = \int \int f\left(x + \frac{x'}{2}, y + \frac{y'}{2}\right) f^*\left(x - \frac{x'}{2}, y - \frac{y'}{2}\right) \times \exp(-i2\pi vx') \exp(-i2\pi \omega y') dx' dy', \quad (3.54)$$

where the x' and y' coordinates are dummy variables used in the integration and (v, ω) are the spatial frequencies corresponding to the (x, y) spatial directions. The transform includes elements of both an auto-correlation and a Fourier transform. As a consequence of keeping both the spatial and spatial-frequency information, the dimensionality doubles from $f \in \mathbb{C}^2$ to $W \in \mathbb{C}^4$. For clarity and out of convention, the remainder of the discussion will look only at the (x, v) components, limiting the Wigner to a two-dimensional function. The arguments can easily be extended to the full 4D case as needed.

Signals which have phase up to quadratic order are represented naturally by Wigner transforms. Consider first a space-domain signal, $f(x)$, written as an envelope function, $A(x)$, multiplied by an additional phase exponential, $\exp(i\phi(x))$. This type of signal could occur in an optical system by inserting a phase transparency as a mask; one common example is a lens. Substituting $f = A \exp i\phi$ into the 1D Wigner equation gives

$$W(x, v) = \int A\left(x + \frac{x'}{2}\right) A^*\left(x - \frac{x'}{2}\right) \exp\left(i\left[\phi\left(x + \frac{x'}{2}\right) - \phi\left(x - \frac{x'}{2}\right) - 2\pi vx'\right]\right) dx'. \quad (3.55)$$

If $\phi(x)$ is a quadratic phase signal, $\phi(x) = \alpha\pi x^2 + 2\beta\pi x + \gamma$. The squared terms in the argument of the exponential cancel yielding a linear argument,

$$\begin{aligned} i \left[\phi \left(x + \frac{x'}{2} \right) - \phi \left(x - \frac{x'}{2} \right) - 2\pi v x' \right] &= i [2\alpha\pi x x' + 2\beta\pi x'] - 2\pi v x' \\ &= -i2\pi [v - \alpha x - \beta] x'. \end{aligned} \quad (3.56)$$

The bracketed term is the frequency argument of the Fourier transform. The first consequence is that a phase offset, γ , does not affect the Wigner transform. The key, however, lies in the induced mapping: what would have appeared at $W\{A(x)\}(x, v) = W_A(x, v)$ is mapped to $W\{f(x)\}(x, v - \alpha x - \beta)$. In other words, the Wigner transform converts multiplicative *quadratic phase* exponentials in the spatial domain into geometric *linear* transforms of the Wigner frequency component.

The Wigner transform can also be defined using Fourier-transformed functions. Taking $F(v) = \mathcal{F}\{f(x)\}(v)$ as the Fourier transform of f , the Wigner transform becomes

$$W(x, v) = \int F \left(v + \frac{v'}{2} \right) F^* \left(v - \frac{v'}{2} \right) \exp(i2\pi x v') dv'. \quad (3.57)$$

Using the same type of arguments as before, operations which result in a quadratic phase exponential in the spatial frequency domain lead to a geometric remapping of the x coordinate of the Wigner transform. One example is that of free-space propagation (see, e.g., the Fresnel kernel, Equation 3.18).

Both mappings can be combined as

$$W\{A\}(x, v) = W\{f\}(ax + bv, cx + dv) = W_f(x', v'), \quad (3.58)$$

for a general optical system with quadratic phase, where here x' and v' represent the remapped variables. (Note that the linear coefficient, β , is set to zero without concern, as most optical components do not shift the coordinate system. An exception would be a glass wedge or prism, but that is not of interest here.) The constants a , b , c , and d relate elegantly to the ABCD matrices [155] of geometric ray optics: as Bastiaans shows, they are the same coefficients [19],[20]. Hence, the Wigner transform provides a powerful link

between geometric and diffractive optics by using the same ABCD ray matrices to remap Wigner coordinates.

Three particular ABCD matrices will be useful here. The first is for free-space propagation,

$$\begin{bmatrix} x' \\ v' \end{bmatrix} = \begin{bmatrix} 1 & \lambda z \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ v \end{bmatrix} = \mathbf{L}_p(z) \begin{bmatrix} x \\ v \end{bmatrix}, \quad (3.59)$$

where z is the propagation distance and λ is the wavelength. The $\mathbf{L}_p(z)$ function represents the propagation matrix. The second is for multiplication by a quadratic phase,

$$\begin{bmatrix} x' \\ v' \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -1/\lambda z & 1 \end{bmatrix} \begin{bmatrix} x \\ v \end{bmatrix} = \mathbf{L}_l(z) \begin{bmatrix} x \\ v \end{bmatrix}, \quad (3.60)$$

where the $\mathbf{L}_l(z)$ function represents a lens-like matrix. The z in this matrix is the axial position at which the quadratic phase appears to converge, which makes sense for both the focal length of a lens and a spherical wave emitted from (or converging towards) a point. Both the \mathbf{L}_p and \mathbf{L}_l matrices result in a shearing transform of the Wigner domain coordinates, the first in the spatial direction and the latter in the spatial-frequency direction. The final ABCD matrix to mention is a magnification,

$$\begin{bmatrix} x' \\ v' \end{bmatrix} = \begin{bmatrix} M & 0 \\ 0 & 1/M \end{bmatrix} \begin{bmatrix} x \\ v \end{bmatrix} = \mathbf{L}_M \begin{bmatrix} x \\ v \end{bmatrix}, \quad (3.61)$$

with M as the magnification factor. This matrix can be derived from the first two by, for example, computing the composite ABCD matrix of a $4f$ system.

3.2.2 Insights to space-bandwidth limits

The geometric remappings of the Wigner transform can be used to understand how the information in a hologram is propagated through a system. The information content of an unknown object is assumed to be contained within a particular space-bandwidth region. While this is never exactly true in reality (a finite spatial extent requires an infinite frequency extent and vice versa), the space-bandwidth region can be arbitrarily set as some fraction

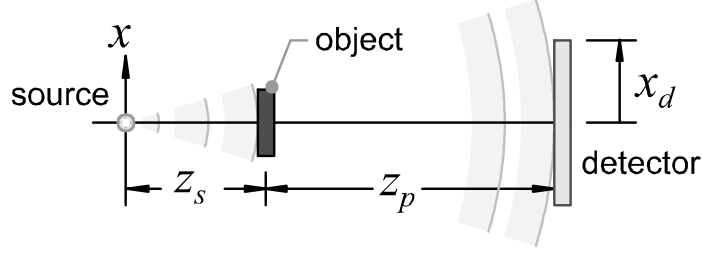


Figure 3-9: Physical variables for describing a spherical reference hologram. The source is a distance z_s from the object, then the reference and object beams propagate a distance z_p to the detector. The detector has half-width x_d .

of the signal's total energy or to contain some components of particular interest to the problem [226]. In keeping with tradition, the object's significant space-bandwidth is denoted in Figure 3-10 as a rectangular region with a dot used as an orientation guide for the eye. Transformations of the Wigner coordinates are visualized as the equivalent transforms applied to the space-bandwidth rectangle.

The most general case for modeling an in-line hologram is using a spherical reference wave, of which a planar reference is the limit as the source distance goes to infinity. The source expands from a real or virtual point source located at a distance z_s from an object of interest, as shown in Figure 3-9. The spherical reference, r , illuminates an object, creating the object wave, o . Both the spherical reference and object waves propagate a distance z_p where they form an interference pattern on a detector. The detector has a half-width of x_d and is able to record spatial frequencies from zero to $\pm v_d = \pm 1/2\Delta$, where Δ is the pixel sampling pitch.

Formation of a hologram in the Wigner domain is shown in Figure 3-10. The initial space-bandwidth of the object is modeled as enclosed within a rectangular region. The product of the spherical reference wave and the object to create the object wave results in a shearing in the v direction, a transformation of (x, v) by $\mathbf{L}_l(-z_s)$. Propagation shears the space-bandwidth region in the x direction, modeled by $\mathbf{L}_p(z_p)$. The useful components of the intensity pattern at the detector are r^*o and o^*r , the first of which is denoted in panel (c). The multiplication of o with r^* is equivalent to transforming the coordinates with $\mathbf{L}_l(z_s + z_p)$. Panel (d) depicts the final state of the object space-bandwidth as recorded on the detector for the o term. The second intensity component, o^*r , is the conjugate, and

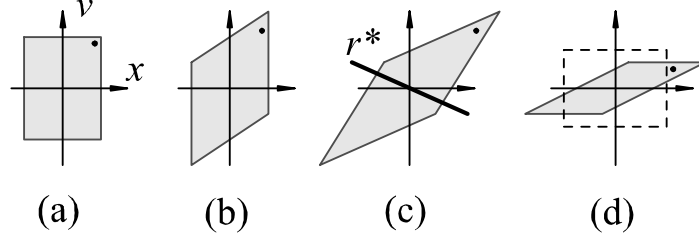


Figure 3-10: Transformations of object bandwidth during hologram formation. The object bandwidth, (a), is sheared vertically in (b) due to the spherical reference wave. Propagation shears the bandwidth horizontally, (c), before the object and conjugate of the reference, r^* , interfere to create the hologram. The final result is depicted in (d), with dotted lines denoting a possible detector bandwidth.

is simply a mirrored copy of r^*o over the $v = 0$ axis due to conjugation properties of the Wigner transform [20],[373].

The composite mapping from a point in the object space, (x, v) , to a point on the detector space, (x', v') , is given by the product of the three ABCD matrices as

$$\begin{aligned}
 \begin{bmatrix} x' \\ v' \end{bmatrix} &= \mathbf{L}_l(z_s + z_p) \mathbf{L}_p(z_p) \mathbf{L}_l(-z_s) \begin{bmatrix} x \\ v \end{bmatrix} \\
 &= \begin{bmatrix} 1 & 0 \\ -1/\lambda(z_s + z_p) & 1 \end{bmatrix} \begin{bmatrix} 1 & \lambda z_p \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1/\lambda z_s & 1 \end{bmatrix} \begin{bmatrix} x \\ v \end{bmatrix} \\
 &= \begin{bmatrix} (z_p + z_s)/z_s & \lambda z_p \\ 0 & z_s/(z_p + z_s) \end{bmatrix} \begin{bmatrix} x \\ v \end{bmatrix} \\
 &= \mathbf{L}_H(z_s, z_p) \begin{bmatrix} x \\ v \end{bmatrix}.
 \end{aligned} \tag{3.62}$$

The matrix $\mathbf{L}_H(z_s, z_p)$ combines the three transform matrices into a holographic system matrix. It can be decomposed to give additional insight into the operation of the entire system [156]. Denoting a magnification factor

$$M = \frac{z_s + z_p}{z_s}, \tag{3.63}$$

\mathbf{L}_H can be written as a product of an \mathbf{L}_M matrix and an \mathbf{L}_p matrix as

$$\mathbf{L}_H = \begin{bmatrix} M & \lambda z_p \\ 0 & 1/M \end{bmatrix} = \begin{bmatrix} 1 & \lambda M z_p \\ 0 & 1 \end{bmatrix} \begin{bmatrix} M & 0 \\ 0 & 1/M \end{bmatrix}. \quad (3.64)$$

The magnification factor is exactly as is predicted from using similar triangles with geometric optics [145]. The propagation matrix from the decomposition 3.64 shows that the magnified object image can be reconstructed by numerically back-propagating using a *planar* reference wave with equivalent distance z_{eq} given by

$$z_{eq} = M z_p = \frac{z_p (z_p + z_s)}{z_s}. \quad (3.65)$$

The results here are in agreement with more laborious analyses [145],[313], but with a significantly streamlined methodology and transparency.

Using \mathbf{L}_H to gain insight to object and detector space-bandwidth limits first warrants a brief divergence into aliasing as it applies to holography. The theory is described by generalized sampling, an extension of Shannon-Nyquist sampling, and accounts for the additional spatial component to the bandwidth of a signal [354],[355]. One-dimensional signal processing shows that sampling a signal with a fixed sampling rate leads to regular replication of the signal spectrum along the frequency direction with spectral replicates centered at the normalized frequencies $\pm 2\pi n$ with n an integer (Figure 3-11). Spectral components beyond normalized frequencies of $\pm\pi$ have the potential to overlap lower frequencies of the neighboring spectral replicates. Those components are ambiguous without *a priori* information about the signal and thus unrecoverable [270].

A similar replication of spectral components occurs when an interference pattern is recorded by a detector. Again, replicates appear centered at $\pm 2n\pi$ in normalized frequency coordinates, and components which extend beyond $\pm\pi$ overlap into neighboring spectral regions. However, consider a signal which has undergone an operation that results in a shift of the spatial coordinate, such as free-space propagation. The components which extend beyond $\pm\pi$ into neighboring replicates do not necessarily overlap the same space-bandwidth region, as shown in Figure 3-11(f). The result is that those aliased components

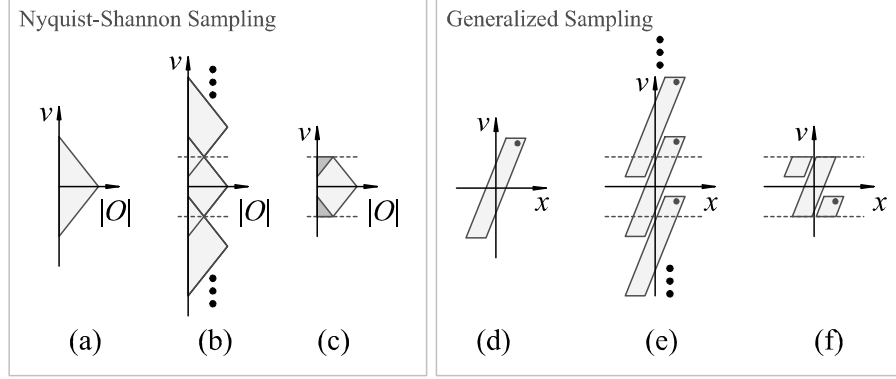


Figure 3-11: Aliased sampling under Nyquist-Shannon and Generalized Sampling. (a), a sample signal spectrum; (b), the discretely sampled spectrum and its virtual replicates; (c), the recorded signal with aliased regions shaded; (d), the space-bandwidth region of a sample signal which has undergone an x -direction shearing operation; (e), the discretely sampled signal and its virtual replicates; (f), the recorded signal. Dotted lines denote the detector Nyquist frequency limit.

can be recovered [356].

The space-bandwidth of the detector used for recording holographic signals can be divided into three sections based on the aliasing regions of generalized sampling, shown in the first panel of Figure 3-12. The first region, denoted by the gray rectangle, extends from $-x_d$ to $+x_d$ in space and $-v_d$ to $+v_d$ in spatial frequency. Signals which fall on this region are recorded without aliasing and can be reconstructed without special processing. The second region extends vertically in the $\pm v$ directions. Holographic signals which fall into this region could be recovered using additional processing if they meet the criteria of generalized sampling [356]. The third region is anything that exists beyond $\pm x_d$. Signals in this region fall off the edge of the detector and are not recorded.

The same procedure that was used to map the object's Wigner transform coordinates into the detector's space-bandwidth can be inverted to determine which components in the object space will reach each detector region. Points (x', v') in the detector space-bandwidth are mapped to

$$\begin{bmatrix} x \\ v \end{bmatrix} = \mathbf{L}_H^{-1}(z_x, z_p) \begin{bmatrix} x' \\ v' \end{bmatrix} = \begin{bmatrix} x'/M - \lambda z_p v' \\ M v' \end{bmatrix}, \quad (3.66)$$

in the object space-bandwidth. The result is a stretch in the v direction and both a shear

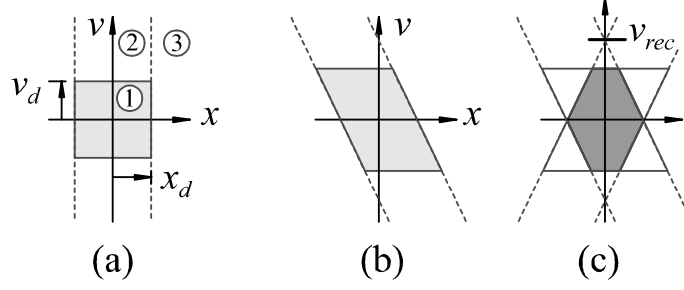


Figure 3-12: Detector bandwidth and its inverse map into the object's Wigner space. A detector space-bandwidth with regions corresponding to (1) normal sampling, (2) generalized sampling, and (3) loss of the signal (a). In (b), the detector space-bandwidth is mapped backwards into the space-bandwidth of the object domain. Part (c) denotes the object bandwidth region corresponding to area (1) for an inline hologram for the case where both o and o^* are recorded.

and a stretch in the x direction. An example of this inverse map is shown in panel (b) of Figure 3-12. Objects whose space-bandwidth information starts in the shaded area will have the matching r^*o intensity term recorded by the detector. The second intensity term, o^*r , which earlier led to a mirroring of the spectral information over the $x = 0$ line, leads to the same effect in the inverse mapping. Thus, to record both the o and o^* components, an object's information must lie in both the shaded region and its mirror over $v = 0$, depicted as the dark shaded area of panel (c).

The inverse mapping of 3.66 can be used to solve for specific space-bandwidth limitations. Consider first an object which extends to $x = x_o$ beyond the optical axis, a detector with half-width x_d , and assume that the methods of generalized sampling hold. The maximum spatial frequency which can be recorded is found by substituting $x = x_o$ into 3.66 and solving for v at the limit, yielding

$$v_{\max} = \frac{1}{\lambda z_p} (x_d - Mx_o). \quad (3.67)$$

As the object's spatial extent x_o shrinks to zero, the maximum recoverable spatial frequency approaches $v_{rec} = x_d/\lambda z_p$, directly related to the idea of the numerical aperture. Frequencies beyond v_{rec} will fall outside the detector region independent of the position of the spherical reference source.

Consider next the case where x_o and the object's maximum useful spatial frequency⁶ v_o are known or can be estimated. This leads to a limit on z_p ,

$$z_p \leq \frac{(x_d - x_o) z_s}{x_o + \lambda z_s v_o}, \quad (3.68)$$

which gives the maximum propagation distance for that object from the detector. A similar type of relation holds for z_s ,

$$z_s \geq \frac{z_p x_o}{x_d - x_o - \lambda z_p v_o}, \quad (3.69)$$

beyond which the signal falls into region 3 (Figure 3-12) and falls outside the detector's spatial region. A common situation is that an object is to be imaged with a desired magnification to meet some resolution criteria, such as for cells on a microscope slide, so that the "best" pair of z_p and z_s are required. The optimal usage of the space-bandwidth of the detector occurs when the equality in 3.68 (or 3.69) holds. Then both z_p and z_s as functions of the magnification, M , can be solved for directly as

$$z_p = \frac{x_d - M x_o}{\lambda v_o}, \quad (3.70)$$

and

$$z_s = \frac{x_d - M x_o}{(M - 1)(\lambda v_o)}. \quad (3.71)$$

The foregoing limits apply in the case when generalized sampling conditions hold so that no restriction was necessary on the maximum spatial frequency in the detector's Wigner space. Including a limit that $v_o/M \leq v_d$, the Nyquist frequency of the detector, gives a limiting condition for the magnification factor,

$$\frac{v_o}{v_d} \leq M \leq \frac{x_d - \lambda z_p v_o}{x_o}. \quad (3.72)$$

Below the minimum magnification set by v_o/v_d , the signal experiences aliasing and generalized sampling is required. Above the maximum magnification, portions of the object signal may miss the detector. Note that equations 3.70 and 3.71 are derived for arbitrary magni-

⁶For example, a 10 μm resolution requires approximately $v_o \simeq 50 \text{ mm}^{-1}$.

fication factors and still hold for M limited by 3.72. Taking the lower limit, $v_o/v_d = M$, gives a condition for the largest non-aliased object extent, the equivalent of half the field of view, as

$$x_{FOV} = x_o \leq \frac{x_d v_d - \lambda z_p v_d}{v_o}, \quad (3.73)$$

which can be readily re-arranged into

$$x_o v_o + \lambda z_p v_o v_d \leq x_d v_d. \quad (3.74)$$

The first term, $x_o v_o$, is the one-quarter the space-bandwidth product (SBP) of the object, while $x_d v_d$ is one-quarter the space-bandwidth product of the detector [226]. (The factor of four is used since both x and v are half-widths.) The conclusion is that, for non-aliased object signals with the largest possible field, the detector must have a SBP not only greater than the object SBP, but greater by $4\lambda z_p v_o v_d = 4\lambda z_p M v_d^2$.

3.2.3 Space-bandwidth product and sampling volume

The choice between detectors for plankton imaging is influenced by the volume of water that they can holographically record. To make the comparison fair, the volume should be computed with the same maximum resolution. This could be achieved by using a set of relay optics to magnify or demagnify an image so that the same target v_o can be recorded. Adding a set of relay optics with magnification M_r results in a coordinate transform given by

$$\begin{bmatrix} x' \\ v' \end{bmatrix} = \begin{bmatrix} M_p M_r & M_r \lambda z_p \\ 0 & 1/M_p M_r \end{bmatrix} \begin{bmatrix} x \\ v \end{bmatrix}, \quad (3.75)$$

where the magnification due to propagation is $M_p = (z_s + z_p)/z_s$, the same as for the case without relay optics. The new system results in a slightly modified set of inequalities for recording a signal without aliasing,

$$\frac{v_o}{v_d} \leq M_p M_r, \quad (3.76)$$

$$M_p M_r x_o + M_r \lambda z_p v_o \leq x_d. \quad (3.77)$$

The first inequality enforces the requirement that v_o is recorded at or below the Nyquist frequency of the detector. The second inequality ensures that object information still reaches the detector. Rearranging the inequalities gives the maximum field of view without aliasing as

$$X_{FOV} = 2 \frac{v_d}{v_o} (x_d - M_r \lambda z_p v_o), \quad (3.78)$$

where X_{FOV} is the total field and is equal to twice the largest x_o . The limit at $X_{FOV} = 0$ gives the maximum propagation distance which will still result in v_o being recorded without aliasing,

$$z_{p,\max} = \frac{x_d}{M_r \lambda v_o}, \quad (3.79)$$

though with the mindful caveat that a zero-width field of view requires an infinitely small object.

The sampling volume without information loss or aliasing is found by integrating the field of view. Defining V_s as the sample volume and assuming a square detector,

$$\begin{aligned} V_s &= \int_0^{z_{p,\max}} X_{FOV}^2 dz_p = \frac{4}{3} \frac{v_d^2 x_d^3}{M_r \lambda v_o^3} \\ &= \frac{N^2 x_d}{12 M_r \lambda v_o^3}, \end{aligned} \quad (3.80)$$

where N is the number of pixels in one direction and N^2 is the full space-bandwidth product of the two-dimensional detector [226]. Since M_r is chosen to maintain the same recordable v_o between detectors, the product $v_d M_r = v_o / M_p = v_{d,eff}$ is a constant. Recalling that $v_d = 1/2\Delta$, where Δ is the detector pixel size, $v_{d,eff} = M_r/2\Delta$. Using the relation $2x_d = N\Delta$, the sampling volume V_s can be simplified:

$$V_s = \frac{N^2}{12 \lambda v_o^3} \frac{(N\Delta/2)}{(2\Delta v_{d,eff})} = \frac{N^3}{48 \lambda v_o^3 v_{d,eff}}. \quad (3.81)$$

Thus, the comparable sampling volume at the same observable resolution is a strong function of the detector space-bandwidth product. A camera with large N is highly preferred over small N , all other properties being similar.

3.2.4 Depth of field

A Wigner viewpoint can also give insight into the apparent depth of field. The depth of field (DOF) has traditionally been defined as $\delta_{DOF} = \lambda/2NA^2$ in a continuous optical system in air, where NA is the numerical aperture [155],[145] (not to be confused with N , the number of samples). In a discrete system, the DOF can be described similarly to the continuous case as the amount of shift in the axial direction which would cause the minimum observable change in the image: one pixel.

The most generic system is that of a spherical wave with relay optics, Equation 3.75. The image can be reconstructed using an equivalent plane wave with $z_{eq} = -M_r^2 M_p$ (consistent with [328] and [313], though without the laborious derivation). The image coordinates, (x_i, v_i) transform as

$$\begin{bmatrix} x_i \\ v_i \end{bmatrix} = \begin{bmatrix} 1 & -M_r^2 M_p \lambda z_p \\ 0 & 1 \end{bmatrix} \begin{bmatrix} M_p M_r & M_r \lambda z_p \\ 0 & 1/M_p M_r \end{bmatrix} \begin{bmatrix} x_o \\ v_o \end{bmatrix} = \begin{bmatrix} M_p M_r x_o \\ v_o/M_p M_r \end{bmatrix}, \quad (3.82)$$

a direct magnification of the original object. Next, consider two point-like objects with near-zero width and some maximum v_o : one in focus, and one that has been displaced axially by δ_z . The in-focus point has information that maps from (x_o, v_o) in the object domain to $(M_p M_r x_o, v_o/M_p M_r)$ in the image domain. The axially displaced object instead has information that maps from $(x_o + \lambda \delta_z v_o, v_o)$ to $(M_p M_r [x_o + \lambda \delta_z v_o], v_o/M_p M_r)$. Since the reconstructions are digital, differentiating between the images would require a one-pixel shift in the image information, so that

$$M_p M_r x_o - M_p M_r [x_o + \lambda \delta_z v_o] = \Delta, \quad (3.83)$$

which gives a simple form for the depth of field as

$$\delta_z = \frac{\Delta}{M_p M_r \lambda v_o} \equiv \delta_{DOF}. \quad (3.84)$$

Equation 3.84 relates directly to the optical DOF quoted in optics textbooks. Take a simple example with $M_p = 1$ and $M_r = 1$. Noting that the detector Nyquist frequency is

$v_d = 1/2\Delta$, δ_z is

$$\delta_z = \frac{1}{2\lambda v_d v_o}. \quad (3.85)$$

Assuming that $v_d \simeq v_o$ and noting that $v_d = \sin \theta / \lambda = NA / \lambda$ for an index of refraction of unity, the DOF is approximately

$$\delta_z \simeq \frac{\lambda}{2NA^2},$$

exactly as expected.

The Wigner formulation of the discrete depth of field, 3.84, makes it immediately obvious that the DOF is directly related to the object spatial frequencies. If $v_o/M < v_d$ in the spherical reference case, then the most limiting factor is not the system NA but the information available from the object, so that the depth of field will change depending on the object itself. This is an effect seen during digital reconstruction, where fine details (with high v) are relatively easy to locate in z , but large or smooth objects (with low v) are difficult to localize axially.

3.2.5 Subtle lessons

The Wigner analysis has a number of subtle lessons applicable to both planar and spherical reference waves that are worth repeating. The assumption here has been that the hologram is in-line; Testorf and Lohmann discuss other lessons for off-axis holograms and generic systems [373], and Stern and Javidi discuss changes for phase shifting [357].

A spherical reference wave is equivalent to a planar reference wave with a depth-dependent magnification term and appropriately adjusted reconstruction distance (Sections 3.2.2 and 3.2.4). Nearly every lesson about planar references can be re-applied to spherical references with only a few changes. One example is the understanding of how motion affects hologram formation⁷ (Section 3.1.1).

Illuminating an object with a spherical reference, $t(x, y) r_s(x, y, z_a)$, pre-distorts its Wigner transform, Figure 3-10(b). The distortion increases at the same rate as the reference

⁷For the inclined reader, modeling motion in the Wigner domain is perhaps more straight-forward as integrating the Wigner along the v direction directly gives the observed intensity. The approach taken here was to model motion using the traditional holography for the sake of consistency with past work – and because working with the Wigner formulation may be new to most readers.

wave as the object information propagates over free space, so that the distortion is exactly cancelled for the r_s^*o and $r_s o^*$ terms, Figure 3-10(c,d). This has a number of consequences. The first is that the reference wave multiplier of o can be seen to cancel out, and the actual intensity information recorded by a hologram is

$$I = |r|^2 + |o|^2 + t_M(x, y) ** h(x, y; z_p) + t_M^*(x, y) ** h^*(x, y; z_p), \quad (3.86)$$

where $t_M(x, y)$ is an appropriately magnified object transparency. (The more traditional $r^*o + ro^*$, Equations 3.11 and 3.13, is still used throughout this work for historical consistency unless noted otherwise.) The same assumptions, of course, still hold: that $|o|^2$ is small compared to the other terms and that $t ** h$ does not significantly distort the reference wave so that an in-line hologram can be formed in the first place. The second consequence is that normalization by the reference, Equation 3.14, is not strictly necessary since the reference has already been cancelled from the intensity terms of Equation 3.86. Normalizing by a spherical reference actually adds an unnecessary phase distortion which must be removed after propagation (see, e.g., [75]).

3.3 DH prototype design

A select number of groups have attempted underwater holography in the last decade, summarized in Tables 3.1, 3.2, and 3.3. Groups led by Watson and Katz both started with film systems, then turned to digital holography around 2004-2007. Owen and Zozulya are credited as having the first marine digital holographic imaging system, targeted specifically at imaging particles, and was notable for illuminating the volume with a diode laser and providing reconstruction software. Kreuzer approaches holography from a physics viewpoint and a strong emphasis on microscopic DH. The work from his group shows an interest in tracking sub-200 μm micro-organisms. Kreuzer et al. are one of the few groups to venture into spherical reference holography. Work from Malkiel, Sheng, and Katz exhibits their interest in behavioral studies of larger plankton, 150–1000 μm . Their DH unit was designed to drift behind a boat with slight positive buoyancy so that it would rise through the water column over a 1-2 hour sampling period, recording *in situ* holographic videos of zooplank-

Device	Investigators	Institution	Target	Status	Refs.
HoloMar	Watson et al.	Aberdeen	plankton	Inactive	[414],[161]
Katz-Film	Malkiel and Katz	Johns Hopkins	plankton	Inactive	[241]
Marine DH	Owen and Zozulya	WPI, Owen Res.	particles	Inactive	[272]
Katz-DH	Malkiel, Sheng, Katz	Johns Hopkins	plankton	Inactive	[281]
eHoloCam	Watson et al.	Aberdeen	plankton	On hiatus	[366],[367]
J/K-DHIM	Kreuzer et al.	Dalhousie Univ.	microplankton	Unknown	[179],[137]
LISST-HOLO	Nimmo Smith	Sequoia Sci.	particles	Commercial	[321],[253]
HoloPOD	Loomis and Davis	MIT-WHOI	plankton	Active	

Table 3.1: Notable holographic devices from the 2000’s. The devices are listed in approximately the order that they appeared in the literature. This thesis is listed as HoloPOD (Holographic Plankton Observation Device; coined by J. Domínguez-Caballero) for comparison.

ton. The eHoloCam developed by Watson et al. received press during 2007-2008 and was used on a limited number of deployments to a few hundred meters. The device has promise as a true ocean-going instrument, though the development appears to have stalled. Finally, the LISST-HOLO is worth mentioning as the only commercial *in situ* DH unit, recently available from Sequoia Scientific. The LISST-HOLO is designed specifically for capturing holograms of particles over long periods of time and comes bundled with a Matlab GUI to allow a user to reconstruct images, then apply their own image processing algorithms to count and size particles.

This section discusses the HoloPOD, a collaboration between MIT and WHOI to create a new holographic imager for real-world plankton studies. The HoloPOD differs from previous devices in its use of a large space-bandwidth product CCD to capture large volumes without avoidance issues and the choice to use a spherical reference for compactness. The remaining part of this chapter gives short overviews of the components, testing, and design that went into the creation of a sea-going prototype unit. Work on the HoloPOD has included a significant component of algorithmic development. Section 3.3.5 briefly introduces software created for this project to interactively reconstruct holograms. Chapter 4 specifically examines methods of retrieving data from holograms with an emphasis on automation – with special attention paid to methods derived from data captured using the HoloPOD prototype.

Device	Imaging vol.	Lat. reso.	SBP	Frame rate	Vol/time	$N^3/\tau \times 10^{-9}$
HoloMar	3500 cc	10 μm	N/A	film	N/A	N/A
Katz-Film	732 cc	3-10 μm	N/A	film	N/A	N/A
Marine DH	10 cc est.	5 μm est.	0.3 MPx	30 fps max	300 cc/s	5.3
Katz-DH	40.5 cc	7.4 μm	4 MPx	15 fps	600 cc/s	129
eHoloCam	36.8 cc	8 μm	6.3 MPx	25 fps max	920 cc/s	425
J/K-DHIM	0.9 cc	1-3 μm	1.4 MPx	7 fps	6.3 cc/s	12.5
LISST-HOLO	1.86 cc	5 μm est.	1.8 MPx	0.2 fps max	0.4 cc/s	0.52
HoloPOD	300 cc	6-14 μm	16-39 MPx	1 fps max	300 cc/s	68-185

Table 3.2: Sampling parameters of the holographic devices. The imaging volume is quoted as per-beam for devices with multiple beams (Katz-DH and HoloMar). The lateral resolution is the maximum quoted by the authors. Volume per unit time is the maximum achievable given the quoted sampling volume and framerate. A better comparison for the detectors is N^3/τ , where τ tau is the period between successive captures; this value is proportional to the sampling volume per unit time for a desired resolution that could be achieved with additional magnification optics.

Device	Reference	Laser	Size	Depth	Power
HoloMar	planar	Q-switched Nd:YAG	$2.4 \times 1 \times 1$ m	100 m	cabled
Katz-Film	planar	Q-switched ruby	$1 \times 1 \times 1$ m	500 m	battery
Marine DH	planar	fiber-coupled diode	$80 \times 20 \times 20$ cm	50 m	cabled
Katz-DH	planar	Q-switched Nd:YLF	$1.2 \times 0.3 \times 0.7$ m	25 m	1 hr. battery
eHoloCam	planar	Q-switched Nd:YAG	$100 \times 30 \times 30$ cm	3000 m	battery
J/K-DHIM	spherical	Nd:YAG	$50 \times 20 \times 20$ cm	20 m	cable
LISST-HOLO	planar	diode	$75 \times 13 \times 13$ cm	300 m	battery
HoloPOD	spherical	fiber-coupled diode	$80 \times 25 \times 30$ cm	1500+ m	battery

Table 3.3: Setups and configurations of the holographic devices. Device size was estimated where not directly available; size includes the mounting platform if necessary for device operation. Depth is the maximum quoted by the authors or, if not given, the operating depth in the references.

3.3.1 System components

Each of the holographic devices in Table 3.1 differs significantly based on their engineering choices and components. This section outlines a few of the key choices made in developing the HoloPOD, both in the laboratory and the prototype. The discussion here is intentionally brief, as most of the engineering choices involve straight-forward calculations [100],[95],[228], though important choices and those which differ from earlier work are paid more attention: in particular, the laser illumination and digital detector.

Many of the early holographic systems use high-power Q-switched lasers (Table 3.3). Pulsed lasers can have extremely high output power and short pulses (on the order of tens of nanoseconds or less). Watson et al. and Katz et al. chose Q-switched lasers for both their film [414],[241] and DH implementations [281],[366],[367],[415] – a fair pairing with CMOS sensors that have low sensitivity compared to CCDs. The downside is that Q-switched lasers tend to be bulky and have greater power requirements [335]. The eHoloCam project specifically collaborated with a laser design firm (Elforlight, Ltd., UK) to re-engineer a suitable Q-switched system for their device [415],[367].

Work with Domínguez-Caballero demonstrated that a high-output single-mode diode laser is sufficient to capture holograms with little to no blurring if a CCD with high sensitivity is used for the detector [100],[95]. A diode laser can be extremely small, on the order of a few centimeters, with the majority of the space occupied by control electronics and packaging. For example, the diodes used for initial laboratory testing were housed in $\varnothing 3.5 \times 15$ cm tubes that included a controller and cooling (PPM modules, Power Technology, Inc.). Later benchtop and device diodes were set into $1 \times 1.2 \times 0.8$ cm mounting blocks (Blue Sky Research, FMXL modules), with $1 \times 1.8 \times 3$ cm controllers (TMD219 driver/controller, Power Technology, Inc.). Laser power ranged from 60 mW to 100 mW.

The lasers were selected to be in the red region of the visible spectrum at 658 nm. This is near the bottom of the absorption curve for seawater (the absorption profile for pure water is plotted in Figure 3-13; Mobley and Aiken et al. note that total absorption is a function of pure water, chlorophyll, and other suspended particles, but that the red absorption is dominated by the pure water component [256],[6]) while remaining well within the operating range of most CCD cameras. A power analysis similar to [95] for the selected

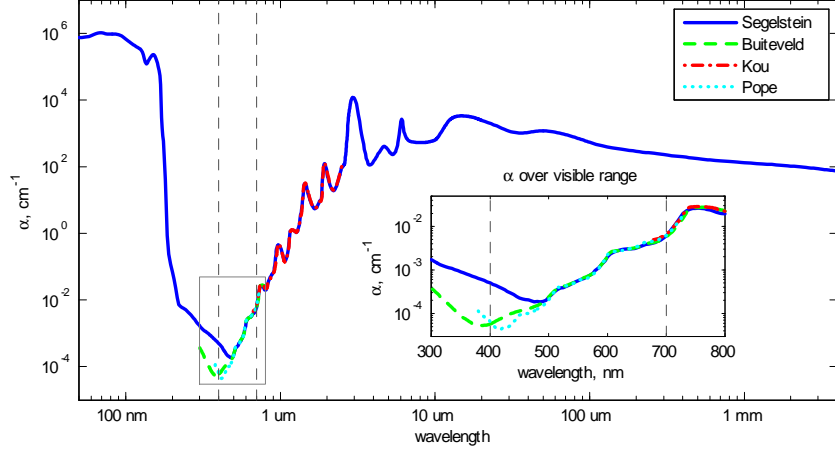


Figure 3-13: Absorption coefficient for light in water, with a minimum around the blue-green region. Data is from [319], [44], [197], and [288]. The curves diverge around 350 - 500 nm due to the extremely low absorption.

cameras and including absorption and scattering [256], Fresnel reflections from windows and lenses (around 4% per air-glass surface unless anti-reflection coated) [155], and losses from spatial filters (measured to be around 40%) indicated that this wavelength would provide enough illumination over a 0.5 m propagation distance. This wavelength has two additional benefits. One is that GaAs-doped diodes, the red and infrared wavelengths, can be produced economically with high outputs while remaining single mode (and thus maintain a high temporal coherence) [335]. The second benefit is that zooplankton have a limited response (i.e., both in the delay before escaping and the magnitude, along with the probability of initiating an escape) to red light, reducing the chance of escape behaviors and avoidance [69],[123],[281],[48].

A mechanical spatial filter was used in the preliminary laboratory experiments and deployments to create a point source with a Gaussian profile. The same effect can be achieved by coupling the laser into a single-mode fiber optic [272],[250],[167] with a short length of fiber required to reduce the high orders [407]. The fiber core has similar dimensions to high NA pinhole filters so that the beam at the exit has the same shape as the previously-tested pinholes, though with a smaller Gaussian width due to the lower NA of the fiber. Fiber-coupled diode lasers from Blue Sky Research (FiberTec II and custom-assembled FMXL modules; 0.13 NA single-mode fiber, 658 nm, 60 mW fiber output) were used for

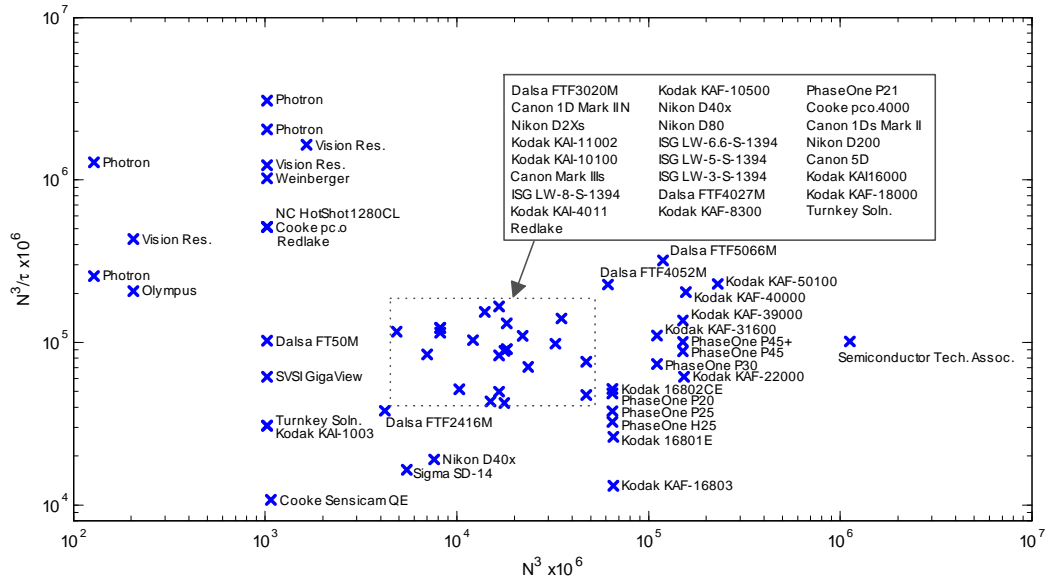


Figure 3-14: Comparative sampling characteristics of select scientific detectors and commercial cameras available during prototype design and construction. Both CCD and CMOS detectors are plotted. A large number of products are available in the middle of the grid, including several commercial CMOS prosumer camera backs. The detector from Semiconductor Technology Associates is a custom-constructed 80 MPx scientific camera for astronomy applications. Detectors to the right of the plot have larger sampling volumes, while detectors towards the top of the plot have greater sampling per unit time.

further laboratory testing, a museum display⁸, two benchtop cruise deployments (Section 3.3.2), and in the prototype unit (Section 3.3.3).

The detector is perhaps the most critical design component and includes a large number of decision variables. It is tied directly to the sampling volume and rate, image quality and resolution, the choice of laser, and the size of the instrument. The choice of a detector depends on the following variables:

- Space-bandwidth product. Equations 3.80 and 3.81 show that the number of pixels

⁸An interactive digital holography museum exhibit was created with Jose A. Domínguez-Caballero and curator Kurt Hasselbalch and installed in the MIT Museum in September 2009. The exhibit uses a Blue Sky fiber-coupled diode laser and a PIC microcontroller for illumination, nearly identical to the components used for the prototype unit (Section 3.3.3), to capture and display holograms at video rates. Users can then pause on a single hologram and explore the volume interactively using custom reconstruction software; that code provided the initial basis for HoloPOD Scientific (Section 3.3.5). As of March 2011, the same Blue Sky fiber-coupled diode laser had provided illumination for around 500 million holograms and more than 100,000 user-selected holograms.

is, to within a set of magnification optics, the deciding factor for the volume that can be imaged at a desired target resolution. Figure 3-14 plots N^3 (proportional to the sampling volume) and N^3/τ (the data rate, proportional to the sampling volume per unit time) for a selection of scientific cameras available as of 2008 (i.e., when the prototype unit was in the design and construction phases). (See also Table 3.2 for comparisons against the previous holographic imagers.) A large sampling volume far from the device is also important to reduce avoidance. Several zooplankton can sense the shear flow around an instrument and will transition to escape behaviors within milliseconds [47],[426],[216],[88],[168]. Reducing the size of the housings and including fairings reduces the shear for a fixed flow speed. However, there will always be a threshold shear near the instrument. A large pixel count extends the potential sampling volume away from the shear region.

- Data rate. The total possible sampling volume is proportional to the data rate, N^3/τ . The CCD architecture in general has slower data rates than CMOS due to the charge transfer and readout method, and the entire chip must be read out for each capture. Larger chips include multiple taps, directing each half (or quarter) of a CCD sensor to different readout electronics and analog-to-digital converters.
- Sensitivity and fill factor. A CMOS chip includes analog-to-digital converters and limited processing electronics in the pixel region, reducing the photosensitive area available for sampling the optical field. A CCD instead uses nearly the entire pixel for sampling, in general leading to significantly better sensitivity and lower noise [175]. High sensitivity enables shorter exposures or a less powerful laser. The lower noise levels in CCDs increases the observed image quality. An interesting note for CMOS is that the small photosite is nearer the ideal point sampling assumed in digital signal processing [270]. Guo et al. address the effect of the fill factor, suggesting that the reconstruction PSF is wider for higher fill factor but has reduced side lobes [150]. In practice, the difference between fill factors is difficult to detect or smaller than other resolution degradations.
- Pixel and detector size. Equation 3.80 gives the maximum sampling volume of a

detector, assuming no aliasing, as

$$V_s = \frac{N^2 x_d}{12 M_r \lambda v_o^3} = \frac{N^3 \Delta}{24 M_r \lambda v_o^3}, \quad (3.87)$$

making the substitution of $2x_d = N\Delta$, where Δ is the pixel size and N is the number of pixels. Systems that are lens-less (i.e., for compactness, simplicity, or robustness) have $M_r = 1$. Thus, the sampling volume is proportional to the pixel size and/or the detector size. It is worth noting that detectors with smaller pixels may have better resolution and will experience aliasing at a further distance, but that trade-off is not enough to increase the overall sampling volume. A larger detector will likely require a larger pressure housing.

- **Implementation.** The interfacing, packaging, power consumption, and cost are all obvious decisions that must be weighed against the design of an instrument. CMOS chips are, in general, less expensive than CCDs and have a lower power consumption [175]. There is also the issue of filters commonly applied over commercial chips: Bayer color filter arrays, infrared filters, and so-called anti-aliasing filters. The issue of signal processing for Bayer filters is addressed in Section 3.3.4. A number of astronomy-minded companies will remove infrared filters from certain models of commercial cameras.

A series of high SBP CCD detectors were tested and utilized in this work. In particular, the high sensitivity of CCDs allowed the use of a diode laser, making the system much more compact and cost-effective. High pixel counts increased sampling volume significantly while reducing the chance of biological avoidance, as discussed above. Specific CCDs and implementations are detailed in the following sections as they were used in practice.

3.3.2 Benchtop systems with oceanic samples

The majority of laboratory test samples came from one of three locations: bottled copepods and diatoms sold in pet stores for establishing an aquarium ecosystem, collected from Boston Harbor, Boston, MA (Long Wharf, 42°21'40.05" N, 71°02'50.73" W) using a plankton net, and from Woods Hole, MA (Eel Pond and WHOI Docks, 41°31'28.15" N, 70°40'12.44" W).

These samples were used to establish the capabilities of a benchtop system and to test the various components of the previous section with similar biologicals as the instrument would image. Initial work on the benchtop system is reported by references [95], [100], and [228].

The benchtop system and components were further tested on two cruise deployments. The samples collected during the cruises differed from the earlier laboratory samples in a number of key aspects:

- The variety of species was markedly different, including a number of large and fragile species. For biologists, it was important to show that the system could faithfully capture images of the different plankton they were interested in studying and to anticipate the visual appearance of these species in the reconstructions. The engineering was benefitted by expanding the test database of holograms to include a wider variety of real-world samples. This is especially important when developing algorithms which rely on an expected size, shape, or appearance. The appearance of trichodesmium colonies and phase objects was particularly illuminating.
- Water properties change significantly between coastal and ocean areas, especially with respect to the concentration of scattering particles [256]. Earlier samples from Boston Harbor were allowed to settle or reconstituted with filtered water in around 80% of the laboratory tests; similarly, bottled samples had to be added to reconstituted seawater to achieve the desired imaging volume for testing. Holograms captured during the cruises focused on including the effects of the media. Unfiltered water was collected at the surface using buckets. Water from depth was sampled from Niskin bottles and specially-designed bottles attached to a remotely operated vehicle.
- The holographic setups were subject to the vibrations of the ships; no motion reduction equipment (elastomer isolators, air dampers, or optical tables) was used. The optical breadboard was secured to the workbench in both cases, further reducing isolation. Despite the range of frequencies emanating from the ship and its engines, imaging resolution measured using a transmissive USAF 1951 plate showed no detrimental effects at the limits of the different setups.

The first cruise traced a path from Panama City, Panama to Woods Hole, Massachusetts

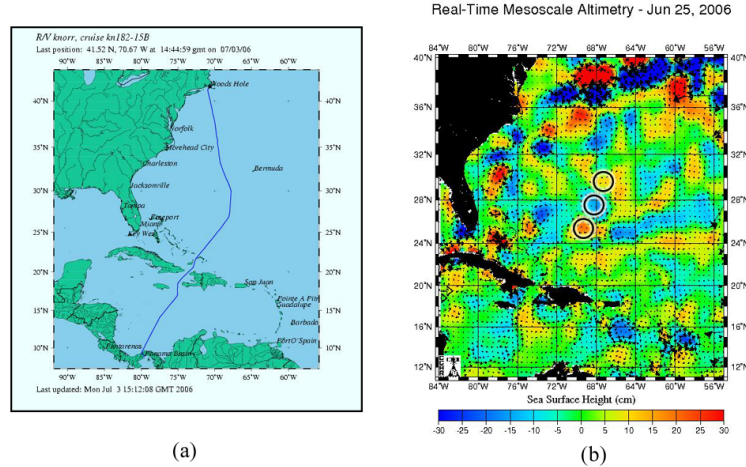


Figure 3-15: Left: Knorr 182-15b cruise path from Panama to Woods Hole, MA, USA through the Caribbean. Right: sea surface height anomalies during the middle of the cruise; circles mark three target vortices where additional samples and cast data were collected.

(Figure 3-15) through the Caribbean Sea during late June of 2006. A benchtop system with a diode laser (Power Technologies, 658 nm, 60 mW), a spatial filter (5 μm pinhole and 60x microscope objective), and a monochrome Kodak KAF-16801E CCD (16 MPx, 9 μm pixel pitch) were used to record holograms. Additional holograms were captured using a fiber-coupled diode laser (Blue Sky Technologies, 658 nm, 100 mW diode output, 60 mW fiber output, single-mode fiber with a 4 μm core). Both planar and spherical reference waves were utilized by either collimating the output from the spatial filter (or optical fiber) or using the raw output. Laser illumination times ranged from 1 μs to 20 μs and were controlled to within 0.5 μs with a PIC (Microchip PIC-16F628A running at 4 MHz). Examples of reconstructed images are shown in Figure 3-16.

The second cruise was in the Celebes Sea in the Philippines during early- to mid-October of 2007. Two benchtop systems were tested. The first matched the benchtop system used aboard the Knorr: a diode laser with a spatial filter used for conditioning and a monochrome KAF-16801E CCD for recording holograms. The second system used components proposed for constructing an *in situ* prototype: a fiber-coupled laser diode, lenses for adjusting the divergence of a spherical reference, and a Kodak KAF-16803C Bayer-patterned CCD sensor (16 MPx, 9 μm pixel pitch). The CCD was embedded into a Hasselblad medium format digital camera back (Hasselblad CFV). Example reconstructions from the Celebes cruise

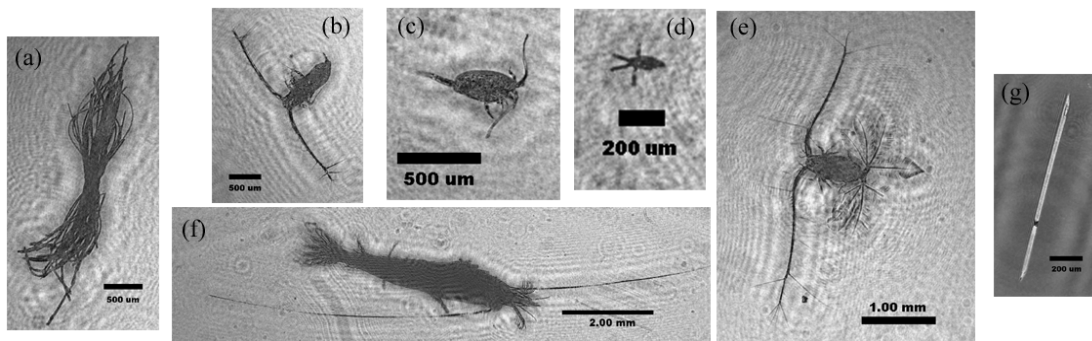


Figure 3-16: Example reconstructions of plankton collected during the Knorr cruise. (a), trichodesmium colony (“bowtie” shape); (b,c,e), three species of copepods; (d) a copepod nauplii; (f) larval lobster; (g) rod diatom.

are shown in Figure 3-17.

The general conclusion from the two cruises was that a fiber-coupled diode laser with the Kodak KAF series CCD chips results in excellent holograms. Around 1600 holograms were captured during the Knorr-Panama cruise and around 2100 while in the Celebes, significantly expanding the database of imaged biologicals, with the caveat that the holograms were all hand-captured during the cruises and thus are more likely to contain objects whose diffraction patterns peaked the interests of the investigators. No effects due to vibration were observed in the images despite the non-ideal working environment, a positive indicator for robustness of the mechanical setups.

3.3.3 Prototype design

A prototype *in situ* holographic imaging device was constructed in early 2008 using lessons learned during laboratory testing and cruise deployments. The device has been used extensively since then to record real-world holograms of plankton, diatoms, oil (Section 4.3), marine snow, particulate matter, and small larvae from various locations around the globe. The main components include a fiber-coupled diode laser for coherent illumination, a lens assembly to create a spherical reference wave, and a high pixel count detector mounted in a commercial camera back to record holograms (Figures 3-18 and 3-19). The prototype was constructed using two housings with an oil-filled connector to communicate between the electronics, contained inside the camera housing, and the laser.

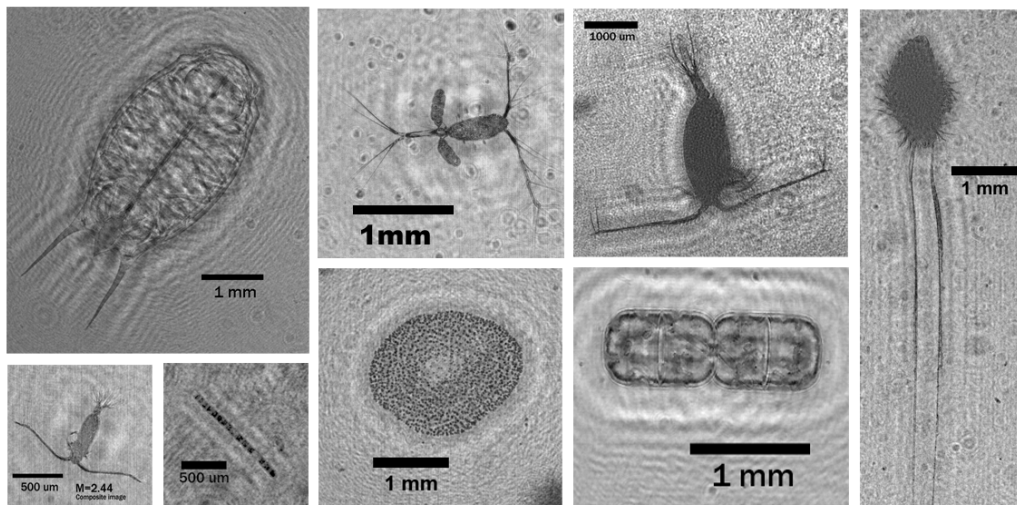


Figure 3-17: Example reconstructions of a salp, copepods, chain diatom, ethmodiscus, bacterial colony, and a ctenophore found during the Celebes cruise.

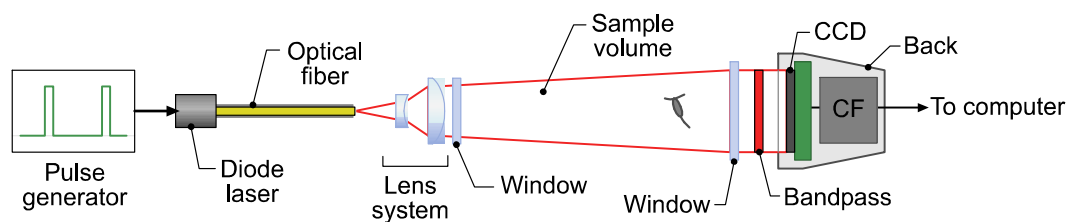


Figure 3-18: Schematic layout of the prototype components.

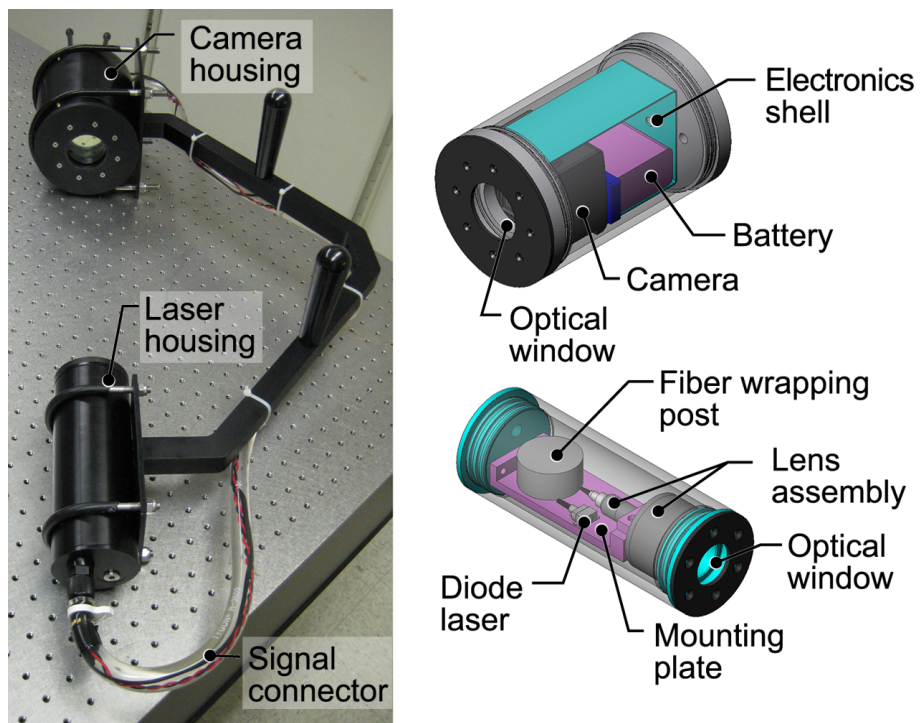


Figure 3-19: Prototype holographic imager, shown mounted to a V-arm with handles. For scale, the holes on the optical table are spaced at 1" intervals. The housing diagrams are courtesy of Sea Scan, Inc.

The prototype was designed with four specific and inter-related goals in mind: a large imaging volume, good image quality, an extended depth rating, and flexible usage:

1. Image volume. Larger separation between the housings and a lower divergence lead to increased sampling volume through increased geometric volume. The total volume imaged at a desired resolution is also proportional to the cube of the number of pixels along the detector's edge, discussed in Section 3.2.3, which is a primary reason for using a camera with high pixel count. The prototype has an imaging volume of around 300-500 mL.
2. Image quality. Identifying plankton based on morphology requires that a good image can be obtained throughout the volume. A camera with high pixel count and low noise is preferred. The ability to adjust the separation between housings is useful to control the number of particles in the volume and thus the amount of cross-talk noise present [96],[98],[97] and Section 3.4. In-line holography provides robust imaging despite vibration and slight misalignment of the optical system (Section 3.1.1). A diode laser with high throughput intensity and a sensitive CCD allow for microsecond illumination periods, reducing motion blur and vibration effects to sub-pixel levels. A lens system is used to expand the spherical reference to around 4 cm at the laser housing's optical window, reducing loss in the quality of the reference from particles near the window. At the same time, the lens system partially collimates the beam to confine the beam's energy to the imaging volume.
3. Depth rating. Evidence suggests that there is microscopic biological activity throughout the water column with surprising amounts below the euphotic zone [184]. A compact and self-contained system makes sampling to greater depths feasible. A diverging spherical reference beam and diode laser allow a smaller housing to be used for illumination. A camera with a large pixel count makes imaging a large volume with good resolution feasible without including magnification optics (see also Section 3.2.3), keeping the camera housing smaller.
4. Flexible usage. The prototype was intended to be an imager which was independent of the carrier platform. The housings have been attached to tether lines, rosettes, sleds,

and remote operated vehicles. The camera and laser systems are controlled internally and data is stored to a memory card in the camera back. The controller can be reprogrammed in a matter of minutes, and camera settings can be adjusted between dives. In-line holography with a spherical reference makes alignment nearly trivial. The laser diode fiber coupling removes operator expertise in adjusting a spatial filter and makes the system more robust. A spherical reference provides a depth-varying magnification (see Section 3.2 and Equation 3.63), so that the system can record images over a wider range of object sizes and resolutions.

A detector with high pixel count, low noise, and good sensitivity is a crucial decision that influences each of the design goals. The two main sensors used in the laboratory benchtop system were a Kodak KAF-16801E (16 MPx monochrome CCD with 9 μm pixels) [107] and a Kodak KAF-16802CE (16 MPx CCD with Bayer color filter array and 9 μm pixels) [108]. The color CCD had lower sensitivity than the monochrome CCD and improved blooming characteristics, but was otherwise similar in performance. A nearly identical color CCD, a Kodak KAF-16803C (16 MPx CCD with Bayer color filter array and 9 μm pixels) [109] was conveniently used in two commercial medium-format camera backs (PhaseOne P20 and Hasselblad CFV) during prototype design and construction in 2007–2008. The backs are compact and self-contained units for capturing and storing images with direct access to the detector chip and facilities for remote triggering, providing a cost-effective solution ideal for the prototype. The CCD in the prototype was replaced in 2010 with a Kodak KAF-39000C (39 MPx CCD with Bayer color filter array and 6.8 μm pixels) [110], embedded in a Hasselblad CFV-39 medium-format camera back.

A bandpass filter reduces the amount of light striking the CCD from wavelengths outside the operating band [366]. This is important for using the imager near the ocean surface where there are significantly higher background light levels and when the imager is mounted on frames with other optically-active devices. The CCD architecture is particularly sensitive to background light and cannot quickly clear the sensor when it has been overexposed. The Hydrogen- α spectral line is conveniently at 656.28 nm, less than two nanometers from the diode laser center wavelength. A number of high-quality $\text{H}\alpha$ bandpass filters are available from astronomical optics suppliers. A 6 nm bandpass filter from Astronomik (Figure 3-

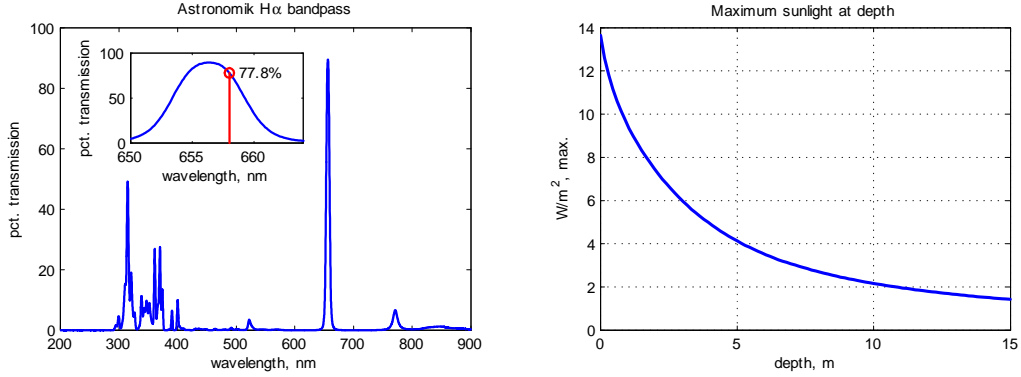


Figure 3-20: Left: H α bandpass filter with 6 nm FWHM passband. The filter has around 78% transmission at the diode laser’s design wavelength. Right: the maximum solar irradiance through the H α filter and a given depth of seawater. The curve is computed as a worst-case scenario: no absorption beyond that of pure water or scattering is assumed, and the receiver surface is oriented with its normal toward the sun. Losses from Fresnel reflections are not included.

20) was selected for the prototype. The filter removes 98.1% of the terrestrial irradiance between 280 and 900 nm reaching the terrestrial surface on a clear day [10] at an average US latitude. Additional absorption and scattering from seawater [256],[6] reduces the light reaching a detector even a few meters below the surface; assuming a worst-case scenario (no scattering or absorption beyond that of pure water), 0.6% of the terrestrial light would pass through 5.0 meters of water and the bandpass filter both. A 36×36 mm detector oriented directly towards the sun receives around 5 mW from the sun at that depth.

The majority of the work in Chapter 4 regarding detection of objects in holograms and focus metrics was done using real-world *in situ* holograms captured by the prototype. Thus, the work in that chapter is directly and immediately applicable to the large quantity of holograms captured using this system.

3.3.4 Power filtering for Bayer detectors

The requirements on detector space-bandwidth, sensitivity, cost, and ease of use may dictate the use of a Bayer-filtered color camera if a suitable monochromatic sensor is not available. The effect of the color filter can be seen by modeling the sensor signal, s , as a series of four

sub-sampled arrays (Figure 3-21),

$$\begin{aligned}
s = & \eta_R o \cdot \text{comb} \left(\frac{x}{2\Delta} - \frac{\Delta}{2}, \frac{y}{2\Delta} + \frac{\Delta}{2} \right) + \eta_B o \cdot \text{comb} \left(\frac{x}{2\Delta} + \frac{\Delta}{2}, \frac{y}{2\Delta} - \frac{\Delta}{2} \right) \\
& + \eta_{G_1} o \cdot \text{comb} \left(\frac{x}{2\Delta} + \frac{\Delta}{2}, \frac{y}{2\Delta} + \frac{\Delta}{2} \right) + \eta_{G_2} o \cdot \text{comb} \left(\frac{x}{2\Delta} - \frac{\Delta}{2}, \frac{y}{2\Delta} - \frac{\Delta}{2} \right) \quad (3.88)
\end{aligned}$$

where o is the original object signal of interest, Δ is the pixel size, $\text{comb}(\cdot, \cdot)$ is the two-dimensional comb function used to model regular sampling, and the η are coefficients for each color channel's sensitivity to the illumination wavelength used to form o . The comb assumes that point sampling at the center of the pixel is sufficient to represent the physics of the sampling; for low spatial frequencies or pixels with small fill factors, this is especially reasonable. (See [203],[204],[150] for further discussion on sampling and fill factors in digital holography.) The green channel is formed by two sub-sampled arrays, denoted as G_1 and G_2 , since in general these may have different hardware or software scaling factors. Taking the Fourier transform of s and applying the shift theorem,

$$\begin{aligned}
S(u, v) = & O(u, v) * [\eta_R C_{2\Delta}(u, v) e^{-i\pi\Delta(-u+v)} + \eta_B C_{2\Delta}(u, v) e^{-i\pi\Delta(u-v)} \\
& + \eta_{G_1} C_{2\Delta}(u, v) e^{-i\pi\Delta(u+v)} + \eta_{G_2} C_{2\Delta}(u, v) e^{-i\pi\Delta(-u-v)}], \quad (3.89)
\end{aligned}$$

where $C_{2\Delta}(u, v) = \mathcal{F}\{\text{comb}(x/2\Delta, y/2\Delta)\}(u, v)$. Applying the scaling theorem in two dimensions,

$$C_{2\Delta}(u, v) = 4\Delta^2 \text{comb}(2\Delta u, 2\Delta v). \quad (3.90)$$

The consequence of convolving $O * C_{2\Delta}$ in each term of 3.89 is to shift the centers of each spectral replicate, normally at multiples of $u = 1/\Delta$, down to $u = 1/2\Delta$ (and similarly for the v direction replicates), shown in Figure 3-22. Since $1/2\Delta$ is the Nyquist frequency corresponding to a $\text{comb}(x/\Delta, y/\Delta)$ sampling, the replicates result in significant high frequency content which, if left unchecked, results in strong noise. The shifted DC components, for example, create a high-frequency checkerboard pattern on reconstructed images.

The effect of the exponentials is easier to see by noting that each $C_{2\Delta}$ is the infinite sum of δ functions at intervals of $1/2\Delta$. The product of the exponentials and $C_{2\Delta}$ then acts to weight each δ by a complex constant. The δ which appears at the origin results in the

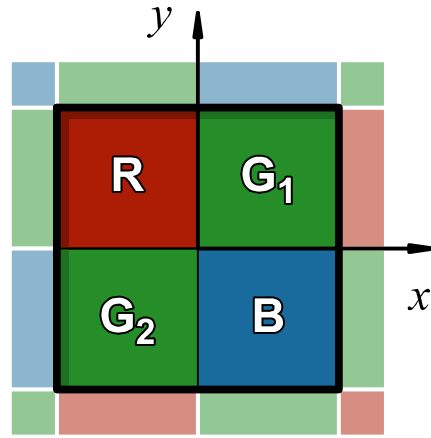


Figure 3-21: Coordinate system depicting a unit cell of a Bayer-patterned color filter. Letters correspond to the color channels.

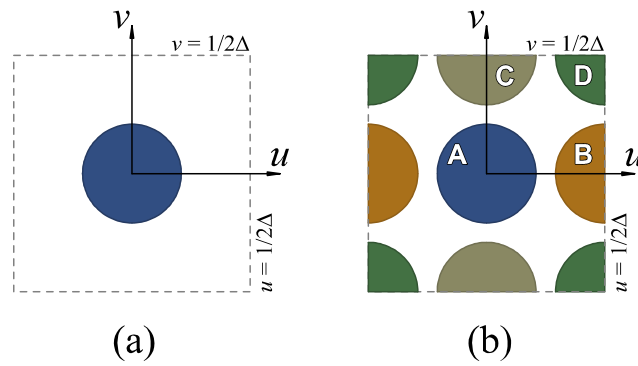


Figure 3-22: Replication of the original spectrum, (a), to the Nyquist frequencies, (b), due to using a detector with a Bayer filter.

spectrum denoted as “A” in Figure 3-22 and is exactly

$$S_A(u, v) = 4\Delta^2 O(u, v) (\eta_R + \eta_B + \eta_{G_1} + \eta_{G_2}) = 4\Delta^2 O(u, v) W_A, \quad (3.91)$$

where W_A acts as a weighting function. Similarly, the other spectra have weights given by

$$W_B = i(\eta_R - \eta_B + \eta_{G_1} - \eta_{G_2}), \quad (3.92)$$

$$W_C = i(-\eta_R + \eta_B - \eta_{G_1} + \eta_{G_2}), \quad (3.93)$$

$$W_D = \eta_R + \eta_B - \eta_{G_1} - \eta_{G_2}. \quad (3.94)$$

There are two interesting conclusions here. One is that if the spectral sensitivity is equal so that $\eta_R = \eta_B = \eta_{G_1} = \eta_{G_2} = \eta$, then all the spectral weights but W_A go to zero. This is the case when a monochromatic sensor is used, or if the color channel gain is adjusted appropriately. (Noise has the possibility of being amplified significantly in the latter case.) The second conclusion is that spectrum “A” can be used directly to reconstruct the object image without needing to know the spectral sensitivities of each component channel.

Filtering spectrum “A” from the other spectral replicates can be performed during the reconstruction step with a suitable window. A windowing filter which has an extended passband and extremely low magnitude at high frequencies is desired. This rules out a majority of the standard filters used in signal processing with the exception of Tukey windows [31],[270]. Even Tukey windows are limited due to their single parameter: the passband can be tuned, but not the high-frequency behavior. Instead, a two-parameter window similar to a Gaussian window is used,

$$P(u) = \exp\left(-\alpha \left|\frac{u}{u_{\max}}\right|^p\right), \quad (3.95)$$

where p is the order of the window P , u_{\max} is the maximum spatial frequency, and α is the negative natural log of the filter’s desired magnitude response at $u = u_{\max}$. The $P(u)$ window, termed a “power” filter⁹, is particularly adept with high p . A $p \geq 2$ is preferred; $p = 2$ gives a Gaussian window [31], $p = 1$ gives a Poisson window [340], and negative p

⁹This filter is referred to as a “super-Gaussian” in beam shaping applications.

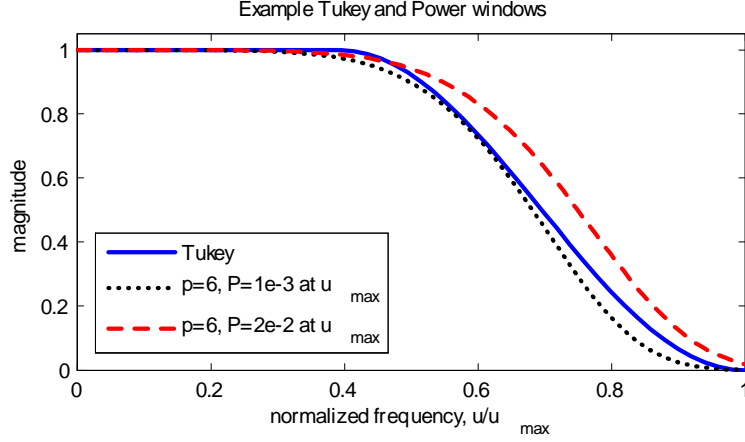


Figure 3-23: Frequency-domain windows: Tukey and $P(u)$, the so-called “power window”.

results in high-pass filters instead. Example one-dimensional realizations of $P(u)$ are shown in Figure 3-23 compared against a Tukey filter of order 0.61.

The width of the power filter in the spatial domain is plotted in Figure 3-24 and can be used to estimate the resolution after applying the power window to an image. The full-width at half-max (FWHM) of the filter response is independent of the image size and is nearly unity over a large range so that the resolution is only slightly worse than with a monochrome sensor and no high-frequency filtering. Typical real-world values of $p \sim 6 - 20$ and $P(u_{\max}) \sim 1 \times 10^{-5} - 1 \times 10^{-1}$ result in a FWHM of $\sim 1.2 - 1.5$, which corresponds well to the experimentally observable resolution of $1.2\Delta - 1.5\Delta$ using resolution targets and the same power filter settings. This FWHM is better than the resolution loss of 2.0 that would be expected from reconstructing images only from the η_R or η_B pixels. Finally, it is worth noting that $p = 2$ gives a traditional exponential window [31] (the far left range of Figure 3-24), which invariably gives significantly increased FWHM and thus extremely poor resolution.

Comparison images of a resolution target reconstructed using various windows, including two power windows, is shown in Figure 3-25. The two power window examples have $p = 4$ and $p = 8$, both with $P(u_{\max}) = 0.1$. The corresponding FWHM values are 1.37 and 1.19 pixels, which corresponds to USAF1951 group 5/element 3 and group 5/element 4, respectively, for this sensor. The zoomed regions show group 5/elements 1-4, and the observed resolution limit is seen to match the FWHM predictions.

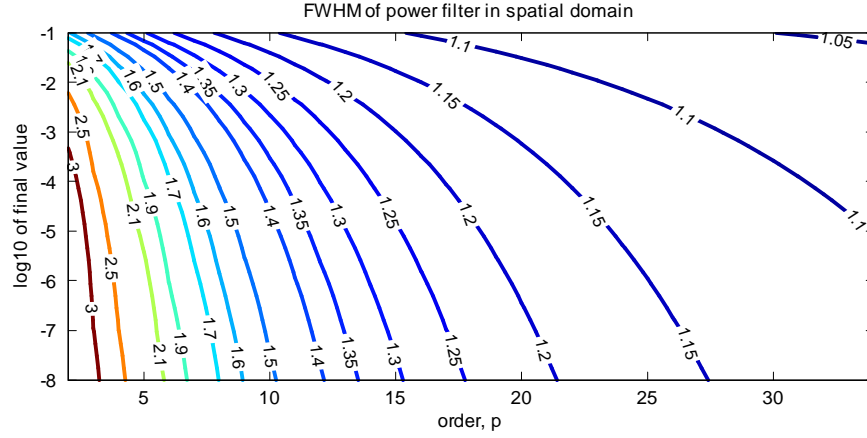


Figure 3-24: FWHM of the power filter's magnitude response in the spatial domain. The final value is defined in the spatial frequency domain as $P(u_{\max})$. The FWHM is in samples (pixels).

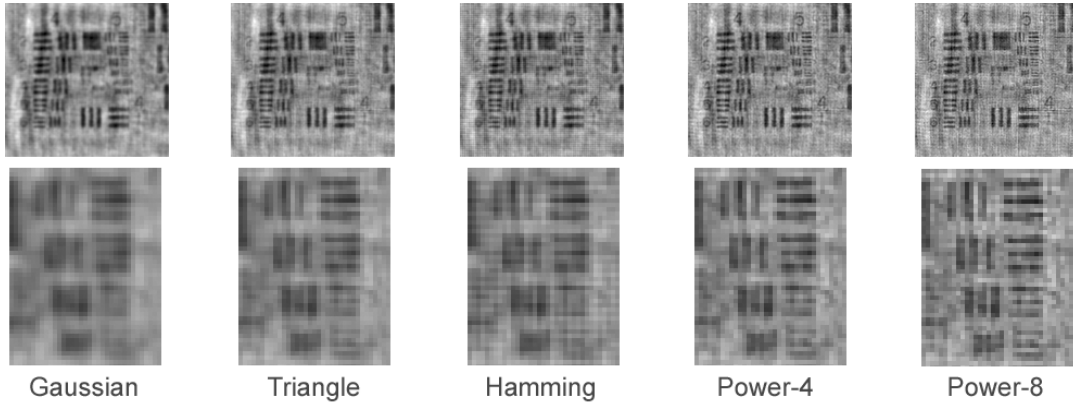


Figure 3-25: Cropped views of windowed reconstructions of USAF1951 resolution targets recorded by a Bayer color filtered sensor. The two power filters have $p = 4$ and $p = 8$ and a final value of $P(u_{\max}) = 0.1$. The zoomed region shows group 5, elements 1-4; the best resolution for this sensor based on the pixel size is group 5, element 5, but is not resolvable with any of the depicted windows.

The alternative to filtering is to apply scaling to each color channel based on measured or estimated spectral sensitivities. The adjusted channels then have equal equivalent η , so that the spectral replicates should cancel and no filtering is necessary. The problem is that noise is also amplified, and the resulting signal-to-noise ratio (SNR) is actually worse. Assume that N is the additive noise and o is again the signal, so that the measured signal is $y_p = \eta_p o_p + N_p$ at pixel p and the original power SNR is $\sum_p \eta_p^2 o_p^2 / \sum_p N_p^2$. Take a simple case of P pixels, equal o_p , and equal N for illustrative purposes (or, equal $\langle o \rangle$ and equal $\langle N \rangle$, more specifically), so that the original SNR is

$$SNR_o = \frac{o^2}{PN^2} \sum_{p=1}^P \eta_p^2 = \frac{\langle \eta^2 \rangle o^2}{N^2}. \quad (3.96)$$

Now consider normalizing the pixels against their spectral sensitivities, so that the measurement becomes $y_p = o_p + N/\eta_p$. The SNR in this scaled case is

$$SNR_{sc} = \frac{Po^2}{N^2 \sum_{p=1}^P \eta_p^{-2}} = \frac{o^2}{N^2 \langle \eta^{-2} \rangle}. \quad (3.97)$$

The SNR in both cases is determined by expectations on η . If η_p is constant for all pixels, the expectations give the same SNR. The difference occurs when η_p varies. Consider a Bayer-patterned sensor with $\eta = \{1, 0.7, 0.7, 0.4\}$ for the four channels. Then $\langle \eta^2 \rangle = 0.535$ and $\langle \eta^{-2} \rangle^{-1} = 0.353$, a 34% loss in SNR. The actual spectral sensitivities at 658 nm for a Kodak KAF-16802CE, the sensor used by PhaseOne and Hasselblad in their 16 MPx camera backs, are closer to $\eta = \{0.18, 0.03, 0.03, 0.02\}$ for the red, green₁, green₂, and blue channels, respectively [108]. These values give $\langle \eta^2 \rangle = 8.65 \times 10^{-3}$ and $\langle \eta^{-2} \rangle^{-1} = 8.42 \times 10^{-4}$ – resulting in a ten-fold loss in power SNR if the color channels are scaled instead of filtered.

3.3.5 Software for interactive reconstructions

Exploring a hologram to understand its contents can be especially challenging for large volumes. Reconstructing slices by hand is extremely time consuming and does not allow for immediate comparison. Writing slices to disk does allow a viewer to compare slices to find, for example, where an object is in focus, but requires a large amount of disk space and can

be slow.

A program, titled HoloPOD Scientific, was created with the goal of interactively reconstructing holograms and saving useful information for later study. Holograms are transferred to a graphics processing unit (GPU) and reconstructed using the Fourier transform method [332],[331]. Keyboard commands or a 3D joystick are used as input to control which slices are reconstructed and at what size. Reconstructions of $4K \times 4K$ pixel holograms are computed at 10-20 fps (faster for smaller reconstructions) and displayed directly to the screen as the real part, imaginary part, magnitude, or phase (Figure 3-26). Individual reconstructions can be saved to disk as either an image or a Matlab file with the complex field data. The power filter can be modified on-the-fly to select the best values given the noise present in each hologram and whether there are spectral replicates due to a Bayer filter (see Section 3.3.4 for more on the power filter).

Along with showing the user reconstructions, HoloPOD Scientific includes the ability to mark objects of interest and to make measurements directly on the screen. The 3D position and reconstruction parameters for each marked object are saved to disk for later use. This allows the user to create a database of points or objects which can be used for machine learning applications and for testing automated algorithms; for example, edge points in real-world holograms were marked using HoloPOD Scientific and used to determine a new set of edge features (Section 4.1.3). Measurements of objects are also written to disk, along with the 3D position and power filter options. The measurements were used to assess the performance of algorithms that estimate object size (Section 4.3) and to create ground truths.

3.4 Particle field effects

One of the largest differences between laboratory experiments and oceanic deployments is the cleanliness of the fluid medium. Oceanic particles, from such sources as bacteria, diatoms, marine snow, silt, mud, and dissolved organic matter permeate real waters. The concentration of these components varies significantly from coastal areas to mid-ocean ranges [256],[157],[324],[300],[5]. High particle concentrations significantly affect the image clarity,



Figure 3-26: Screenshot of HoloPOD Scientific in use. The hologram is shown in the right panel, with the blue square marking pixel area used for reconstructions and the embedded green square marking the area which is being viewed by the user. The left panel shows the magnitude of the reconstructed field, zoomed in to the barnacle nauplii. A plot of the power filter magnitude and the reconstruction parameters are displayed below the reconstruction image. Users can mark or measure objects of interest by clicking within the reconstruction area and the positions are automatically saved to disk.

limiting the volume of water that can be imaged. This section aims to predict the effects of particles on the hologram spectrum, providing intuition about how particles might degrade the reconstructions.

The effect of particles is two-fold: scattering and absorption from particle fields provides both noise and a reduction in the desired signal. The noise, considered here to be the hologram signal due to random particles, can be estimated by finding the expected spectrum given some size distribution. A number of additional assumptions are necessary:

1. Particles can be modeled as having an opacity of s ($s = 0$ is transparent, $s = 1$ is opaque). Most observed particles have a high enough phase variation, either through surface and internal roughness or index of refraction difference from the surrounding medium, that they can be considered nearly opaque for digital holographic measurements if they are not already opaque.
2. The probability of finding a particle at a particular 3D location is not dependent on the other particles present. In other words, the number and size of particles is low enough that there are not significant interactions either in occupation of space or from fluid dynamics.
3. The expected number of particles in the observation volume, N , is known or can be estimated. For example, Beer's law can be shown to hold for particles in both planar and spherical reference configurations, so that $\langle Ns \rangle$ can be estimated from the hologram intensity if the water's absorption is known. Estimation can also be done using instruments such as the popular WET Labs AC-9 spectrophotometer [418] (see, e.g., [15] for use of the b coefficient to estimate distribution characteristics or [300] for estimation from the beam attenuation coefficient) or Secchi disks [424]. Direct measurement of large particles can be done using digital holography (see discussion of oil droplet size distribution estimation in Section 4.3), or of small particles using laser scatterometers such as Sequoia Scientific's LISST [320] or Coulter Counters (e.g., [23]) [300].
4. The amount of re-scattering of one particle's object wave from another is assumed to be small, so that the hologram is treated as a sum of object waves from each individual

particle.

The intensity at the hologram plane is

$$I(x, y) = \left| r - \sum_j s_j t_j(x, y) * h_f(x, y; z_j) \right|^2, \quad (3.98)$$

$$= |r|^2 - r^* \sum_j o_j - r \sum_j o_j^* + 2 \sum_j \sum_k o_j o_k^*, \quad (3.99)$$

where $t_j(x, y)$ is the particle transparency function of the j^{th} particle, s_j is the opacity, $h_f(x, y; z_j)$ is the standard Fresnel kernel (Equation 3.18), and $o_j = s_j t_j * h_f(z_j)$ is the resulting object wave for the j^{th} particle. The Fresnel kernel is preferred here over the exact kernel for its mathematical properties. This is appropriate given the high variability and uncertainty in real-world particles, so that the minor loss of accuracy from the substitution is trivial in comparison. The negative sign before the sum denotes that the particle is treated as absorbing and thus the real part of t_j is positive. Using the traditional assumptions that r is a plane wave with unitary magnitude and $|o_j \sum_k o_k^*| \ll |o_j r^*|$ so that the cross-talk terms can be considered negligible background, 3.99 becomes

$$I(x, y) = 1 - \sum_j (o_j + o_j^*). \quad (3.100)$$

The Fourier transform of the intensity is

$$\widehat{I}(u, v) = \delta(u, v) - \sum_j [O_j(u, v) + O_j^c(u, v)], \quad (3.101)$$

using

$$\mathcal{F}\{o_j(x, y)\} = O_j(u, v) = s_j T_j(u, v) H_f(u, v; z_j) \exp(-i2\pi[uq_x + vq_y]), \quad (3.102)$$

and

$$\begin{aligned}\mathcal{F}\{o_j^*(x, y)\} &= O_j^c(u, v) = s_j T_j^*(-u, -v) H_f^*(-u, -v; z_j) \\ &\times \exp(-i2\pi[uq_x + vq_y]),\end{aligned}\quad (3.103)$$

as the Fourier transforms of the object waves; T and $H_f(z_j)$ are the Fourier transforms of t_j and $h_f(z_j)$. Here, the particle transparency mask is defined at the origin and laterally shifted to a center at (q_x, q_y) .

The restriction on r can actually be relaxed to say that r is assumed to *behave* like a plane wave, as the Wigner approach and Equation 3.64 showed that spherical reference waves are indistinguishable from planar references at the hologram plane aside from their induced geometric magnification of the object. Thus, the derivation here extends to spherical waves easily by scaling t_j laterally and replacing z_j with $z_{eq,j}$. The specifics of a spherical reference will be discussed later in this section.

The spectral energy gives the expected signal power at each frequency component, calculated as

$$\Psi = \left\langle \left| \hat{I}(u, v) \right| \right\rangle = \left\langle \hat{I}(u, v) \hat{I}^*(u, v) \right\rangle. \quad (3.104)$$

Substituting 3.101 yields

$$\begin{aligned}\Psi &= \left\langle \delta - \delta \sum_j (O_j + O_j^c + O_j^* + O_j^{c*}) + \sum_j \sum_k O_j O_k^* + O_j^c O_k^* + O_j O_k^{c*} + O_j^c O_k^{c*} \right\rangle, \\ &= \delta - \delta N \langle O + O^* + O^c + O^{c*} \rangle\end{aligned}\quad (3.105)$$

$$+ \left\langle \sum_j \sum_k O_j O_k^* + O_j^c O_k^* + O_j O_k^{c*} + O_j^c O_k^{c*} \right\rangle, \quad (3.106)$$

for the spectral power. The first term is the power of the normalized reference. The second term needs only to be evaluated at $(u, v) = (0, 0)$ due to the multiplication by the δ function and can be rewritten as

$$\delta N \langle O + O^* + O^c + O^{c*} \rangle = \delta N [\langle O \rangle + \langle O^* \rangle + \langle O^c \rangle + \langle O^{c*} \rangle]. \quad (3.107)$$

Computing the expectation for the O component,

$$\langle O(0,0) \rangle = \int p(s, z, \boldsymbol{\theta}, q_x, q_y) s T(0, 0; \boldsymbol{\theta}) H_f(0, 0; z) ds dz d\boldsymbol{\theta} dq_x dq_y, \quad (3.108)$$

where $\boldsymbol{\theta}$ is treated as a set of parameters which describe the possible masks. The joint distribution $p(s, z, \theta, q_x, q_y)$ is not strictly necessary, as each variable is treated as independent under the problem assumptions and the functions are separable. The expectation becomes

$$\langle O(0,0) \rangle = \int p(s) s ds \int p(z) dz \int p(\boldsymbol{\theta}) T(0, 0; \boldsymbol{\theta}) d\boldsymbol{\theta} \quad (3.109)$$

$$\times \int p(q_x) dq_x \int p(q_y) dq_y \quad (3.110)$$

$$= \langle s \rangle \int p(\boldsymbol{\theta}) T(0, 0; \boldsymbol{\theta}) d\boldsymbol{\theta}, \quad (3.111)$$

and similarly for the other $\langle O \rangle$ -type expectations. Noting that $T(0,0)$ is related to the cross-sectional area of the particles, the second term of Equation 3.106 can be seen to represent the power lost to absorption.

The double sums in Equation 3.106 can be written as self- and cross- terms. Taking the first component as an example,

$$\left\langle \sum_j \sum_k O_j O_k^* \right\rangle = \left\langle \sum_j O_j O_j^* \right\rangle + \left\langle \sum_j \sum_{k, k \neq j} O_j O_k^* \right\rangle, \quad (3.112)$$

$$= N \langle O O^* \rangle + N(N-1) \langle O \rangle \langle O^* \rangle, \quad (3.113)$$

which assumes that O_j and O_k^* are sufficiently independent. The $\langle O \rangle$ expectation is

$$\begin{aligned} \langle O \rangle &= \int \cdots \int p(s, z, \boldsymbol{\theta}, q_x, q_y) s T(u, v; \boldsymbol{\theta}) H_f(u, v; z) \\ &\quad \times \exp(-i2\pi[uq_x + vq_y]) ds dz d\boldsymbol{\theta} dq_x dq_y \end{aligned} \quad (3.114)$$

$$\begin{aligned} &= \int p(s) s ds \int H_f(u, v; z) p(z) dz \int p(\boldsymbol{\theta}) T(u, v; \boldsymbol{\theta}) d\boldsymbol{\theta} \\ &\quad \times \int \exp(-i2\pi u q_x) p(q_x) dq_x \int \exp(-i2\pi v q_y) p(q_y) dq_y, \end{aligned} \quad (3.115)$$

again assuming that $s, z, \boldsymbol{\theta}, q_x$ and q_y are independent. Taking the maximum beam lateral extents to be $\pm x_d$ and $\pm y_d$, and assuming the particle positions are uniformly distributed

over the beam, $p(q_x) = 1/2x_d$ and $p(q_y) = 1/2y_d$. The q_x integral evaluates to

$$\int p(q_x) e^{-i2\pi u q_x} dq_x = \frac{1}{2x_d} \int_{-x_d}^{x_d} e^{-i2\pi u q_x} dq_x, \quad (3.116)$$

$$= \frac{\sin(2\pi u x_d)}{2\pi u x_d} = \text{sinc}(2u x_d), \quad (3.117)$$

and similarly for q_y . The sinc argument can be simplified by assuming that x_d is the half-width of the detector (or on a similar order) and noting that the maximum u is the Nyquist frequency of the detector, $u_{Nyq} = 1/2\Delta = n_{pix}/4x_d$, where n_{pix} is the number of pixels in the detector. Substituting yields

$$\text{sinc}(2u x_d) = \text{sinc}\left(\frac{1}{2} \frac{u}{u_{Nyq}} n_{pix}\right),$$

which decays quickly, affecting only the lowest spatial frequencies. Assuming that the particles are uniformly distributed over $z = z_a$ to $z = z_b$, then $p(z) = 1/(z_b - z_a)$ and the remainder of the $\langle O \rangle$ integration evaluates to

$$\begin{aligned} \langle O \rangle &= \langle s \rangle \text{sinc}(2u x_d) \text{sinc}(2v y_d) \int p(\boldsymbol{\theta}) T(u, v; \boldsymbol{\theta}) d\boldsymbol{\theta} \\ &\times \frac{\exp(-i\pi\lambda z_b \rho^2) - \exp(-i\pi\lambda z_a \rho^2)}{-i\pi\lambda \rho^2 (z_b - z_a)}, \end{aligned} \quad (3.118)$$

where $\rho^2 = u^2 + v^2$.

Equation 3.118 shows that $\langle O \rangle$ decays like $1/u^3$ from the combination of the $1/\rho^2$ and sinc terms, so that the product $\langle O \rangle \langle O^* \rangle$ in 3.113 decays like $1/u^6$. For reasonable particle densities, the $\langle O \rangle \langle O^* \rangle$ (and similar terms) contribute a small amount to extremely low spatial frequencies, but can safely be ignored for mid- to high spatial frequencies. Removing these cross terms from 3.106, the spectral power is approximately

$$\Psi \simeq \delta - \Upsilon(s, \boldsymbol{\theta}) + N \langle OO^* \rangle + N \langle O^c O^* \rangle + N \langle OO^{c*} \rangle + N \langle O^c O^{c*} \rangle, \quad (3.119)$$

where $\Upsilon(s, \boldsymbol{\theta})$ represents the absorption loss and affects only the DC component.

The $\langle OO^* \rangle$ and similar terms of Equation 3.119 are where the interesting relationships

are hidden. Computation of the expectation,

$$\langle OO^* \rangle = \int p(s) s^2 ds \int p(z) dz \int p(q_x, q_y) dq_x dq_y \int p(\boldsymbol{\theta}) T(u, v; \boldsymbol{\theta}) T^*(u, v; \boldsymbol{\theta}) d\boldsymbol{\theta}, \quad (3.120)$$

is aided by the fact that $H_f H_f^* = 1$ and the lateral shifts cancel, so that

$$\langle OO^* \rangle = \langle s^2 \rangle \int p(\boldsymbol{\theta}) T(u, v; \boldsymbol{\theta}) T^*(u, v; \boldsymbol{\theta}) d\boldsymbol{\theta}. \quad (3.121)$$

A similar set of results for $\langle O^c O^{c*} \rangle$ gives

$$\langle O^c O^{c*} \rangle = \langle s^2 \rangle \int p(\boldsymbol{\theta}) T(-u, -v; \boldsymbol{\theta}) T^*(-u, -v; \boldsymbol{\theta}) d\boldsymbol{\theta}. \quad (3.122)$$

The $\langle O^c O^* \rangle$ and $\langle OO^{c*} \rangle$ terms are fundamentally different in how the H_f kernel is treated. Taking the first of the two terms,

$$\begin{aligned} \langle O^c O^* \rangle &= \int p(s) s^2 ds \int p(q_x, q_y) dq_x dq_y \int p(\boldsymbol{\theta}) T^*(-u, -v; \boldsymbol{\theta}) T^*(u, v; \boldsymbol{\theta}) d\boldsymbol{\theta} \\ &\times \int_{z_a}^{z_b} p(z) e^{i2\pi\lambda z \rho^2} dz. \end{aligned} \quad (3.123)$$

The matched conjugation of the kernel in O^c and O^* results in a factor of two in the exponentials. Evaluating the integrals gives

$$\begin{aligned} \langle O^c O^* \rangle &= \frac{\langle s^2 \rangle}{(z_b - z_a)} \int p(\boldsymbol{\theta}) T^*(-u, -v; \boldsymbol{\theta}) T^*(u, v; \boldsymbol{\theta}) d\boldsymbol{\theta} \\ &\times \frac{\exp(i2z_b\pi\lambda\rho) - \exp(i2z_a\pi\lambda\rho)}{i2\pi\lambda\rho^2}. \end{aligned} \quad (3.124)$$

Similarly, $\langle OO^{c*} \rangle$ results in

$$\begin{aligned} \langle OO^{c*} \rangle &= \frac{\langle s^2 \rangle}{(z_b - z_a)} \int p(\boldsymbol{\theta}) T(u, v; \boldsymbol{\theta}) T(-u, -v; \boldsymbol{\theta}) d\boldsymbol{\theta} \\ &\times \frac{-\exp(-i2z_b\pi\lambda\rho) + \exp(-i2z_a\pi\lambda\rho)}{i2\pi\lambda\rho^2}. \end{aligned} \quad (3.125)$$

Note here the modification to the T function and its arguments as well.

The expectations can be combined if T is real and even or can be modelled as such.

Examples include round objects such as droplets and bubbles, or diatoms and bacteria which scatter like circular objects within the frequency bands of interest – as is often the case within the low spatial-frequencies recorded by digital holography. Combining and simplifying the self-power terms under the assumption that T is real and even,

$$\Psi_S \equiv N [\langle OO^* \rangle + \langle O^c O^{c*} \rangle + \langle O^c O^* \rangle + \langle OO^{c*} \rangle], \quad (3.126)$$

$$\begin{aligned} &= N \langle s^2 \rangle \int p(\boldsymbol{\theta}) T^2(u, v; \boldsymbol{\theta}) d\boldsymbol{\theta} \\ &\quad \times \left[2 + \frac{\sin(2z_b \pi \lambda \rho^2) - \sin(2z_a \pi \lambda \rho^2)}{(z_b - z_a) \pi \lambda \rho^2} \right], \end{aligned} \quad (3.127)$$

or

$$\Psi_S = 2N \langle s^2 \rangle [1 + \text{sinc}(z_\Delta \lambda \rho^2) \cos(2\bar{z} \pi \lambda \rho^2)] \int p(\boldsymbol{\theta}) T^2(u, v; \boldsymbol{\theta}) d\boldsymbol{\theta}, \quad (3.128)$$

$$= 2N \langle s^2 \rangle [1 + \text{sinc}(z_\Delta \lambda \rho^2) \cos(2\bar{z} \pi \lambda \rho^2)] \langle T^2 \rangle, \quad (3.129)$$

using $z_\Delta = z_b - z_a$ as the depth range and $\bar{z} = (z_a + z_b)/2$ as the mean depth. Equation 3.128 implies that the spectral energy from the self-terms should have a spatial-frequency decay rate that goes with z_Δ and limits to a constant, non-zero value as $\rho \rightarrow \infty$.

A set of simulations showed that Ψ_S was the dominant component of Ψ . A volume of particles was constructed by approximating the volume as a series of slices. Circular particles were placed on each slice using uniform distributions for the lateral positions and a Poisson distribution for the total number of particles expected to be within the volume represented by each particular slice. Combinations of discrete mono-disperse distributions were used for the circular particle size parameter $\boldsymbol{\theta}$. The field between each slice was propagated using the higher-accuracy exact propagation kernel, Equation 3.17. Additional accuracy for small particles was ensured by using anti-aliased particle images and oversampling in the simulation domain. An example hologram is shown in Figure 3-27. The spectral energy was calculated by summing the energy of the discrete Fourier transforms of 1250 realizations of particle fields generated in this manner. The spectral energy and its frequency-dependent components are shown in Figure 3-28. The analytical solution is shown for Ψ_S using the assumptions discussed in this section, and notably does not include the $\langle O \rangle \langle O^* \rangle$ -type terms,

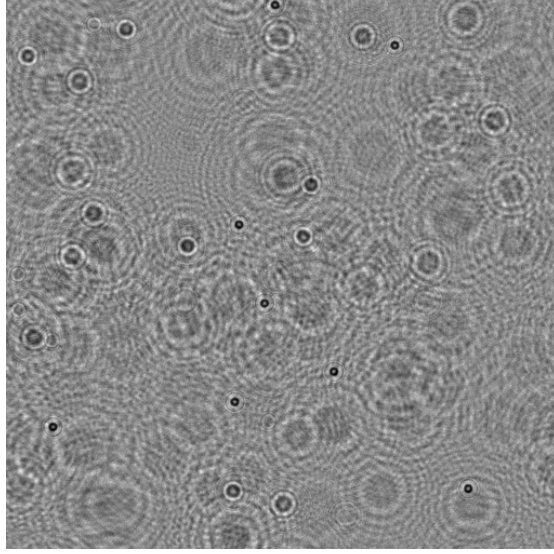


Figure 3-27: Sample simulated particle hologram with two mono-disperse size distributions which have equal probability.

the absorptive terms, multiple diffraction (which is included in the simulation), or the terms which would result from including the $|o|^2$ -like “halo” terms in the model. Despite this, Ψ_S represents the spectral energy well.

The Ψ_S of Equation 3.128 includes an integral over $T(u, v; \boldsymbol{\theta})$, a function of the particle shape and size. A number of particle size distributions have been proposed for oceanic particles [330]. The most commonly used form is the Junge-type or power law distribution,

$$N'(D) = k \left(\frac{D}{D_o} \right)^{-\gamma}, \quad (3.130)$$

where D_o is a reference diameter, k is the differential concentration at D_o , and γ typically lies between 3 and 4 but has a wide distribution that ranges from around 2 to 4.5 [300],[256],[330]. The associated $p(\boldsymbol{\theta})$ is given by normalizing $N'(D)$,

$$p(a) = \frac{\gamma - 1}{a_m^{1-\gamma}} a^{-\gamma}, \quad (3.131)$$

where a is the particle radius and a_m is the minimum particle considered. If the particles

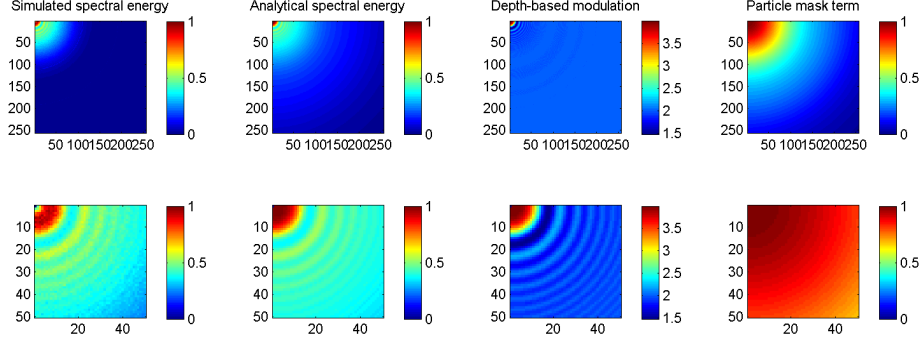


Figure 3-28: Simulated and analytic spectral energy. From left to right: simulation, analytic approximation using Ψ_S , the component of Ψ_S from the depth integral, and the component of Ψ_S from the particle mask integral. The top row are the spectral components in a single quadrant, while the bottom row are zoomed views of the same component. Spectral energy plots were normalized to their maximum values for comparison. Grid points are frequency samples, with the Nyquist limit at sample 256.

are assumed to have circular cross-sections with radius a , $T(a)$ is given by

$$T(\rho; a) = \frac{a J_1(2\pi a \rho)}{\rho} = a\pi \text{jinc}(a\rho), \quad (3.132)$$

where J_ν is a Bessel function of the first kind [145]. Substituting Equations 3.131 and 3.132, the $\langle T^2 \rangle$ integral is

$$\langle T^2 \rangle = \int_{a_m}^{\infty} \frac{\gamma - 1}{a_m^{1-\gamma}} a^{-\gamma} [a\pi \text{jinc}(a\rho)]^2 da, \quad (3.133)$$

$$= \frac{K_1(\gamma, a_m)}{\rho^{5-\gamma}} [K_2(\gamma) + K_3(\gamma, a_m) \rho^{5-\gamma} F_{2,3}(-a_m^2 \rho^2)], \quad (3.134)$$

where $F_{p,q}$ is a generalized hypergeometric function which decays quickly and the K_i are constants. The result is that $\langle T^2 \rangle$ is strongly dominated by its $\rho^{\gamma-5}$ term, demonstrated by nearly straight lines on a log-log plot (Figure 3-29). Variations due to the Bessel function are smoothed out.

What is particularly interesting is what happens at a typical value of $\gamma = 3$. The resulting $\langle T^2 \rangle$ spectra are nearly matched in slope to ρ^{-2} (plotted for comparison as the solid black line with arbitrary offset), which is the expected decay rate of power spectra for natural images and scenes (see, e.g., [397],[16],[386] and references therein). A γ of

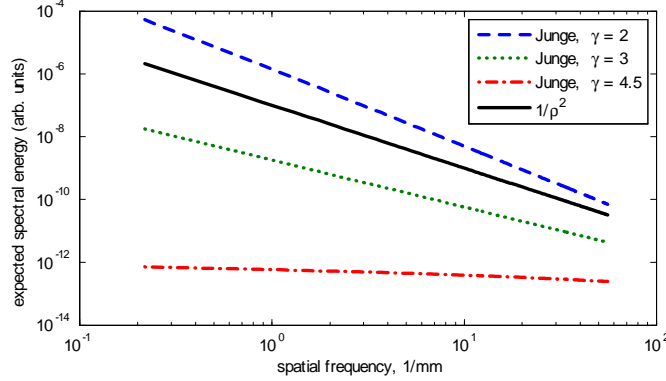


Figure 3-29: Expected spectra for a Junge-type distribution of circular particles.

around 2 – 3 better corresponds to the slightly higher decay rate observed in underwater images [16]. The fact that $\langle T^2 \rangle$ has a ρ^{-a} relationship which matches previous research is particularly encouraging considering the assumptions and simplifications that were used in the derivation.

Sample plots of other popular particle size distribution models [330] and the corresponding $\langle T^2 \rangle$ are depicted in Figure 3-30. Curiously, the log-normal, beta, and gamma distributions all have slopes of -3 in the log-log plot, as would a Junge-type distribution with $\gamma = 2$. The distributions show strong similarities in their spectral shape and slope, especially for moderate to high spatial frequencies. The agreement suggests that the model is robust to the specific $T(\theta)$ or particle size distribution selected, so that a generic ρ^{-a} model for the spectral energy from a collection of particles is reasonable. This also explains why estimating the full particle size distribution from the spectra may be especially difficult without distinctive distributions; as of writing, only the mean particle size has been approached [92]¹⁰.

The specific ρ^{-2} model describes observed noise in “natural holograms” (i.e., those without man-made objects such as resolution targets) extremely well. An ensemble of 1,621 holograms captured in lab and field settings was used to estimate the spectral energy of natural holograms. Each hologram was recorded by a monochrome detector with 4096 by 4096 pixels. The spectral energy was computed for each hologram, normalized to the same

¹⁰The method proposed by [92] uses the peak around $(u = 0, v = 0)$, the area that has slightly greater discriminability for the different distributions – and contributions from the $\langle O \rangle \langle O^* \rangle$ terms.

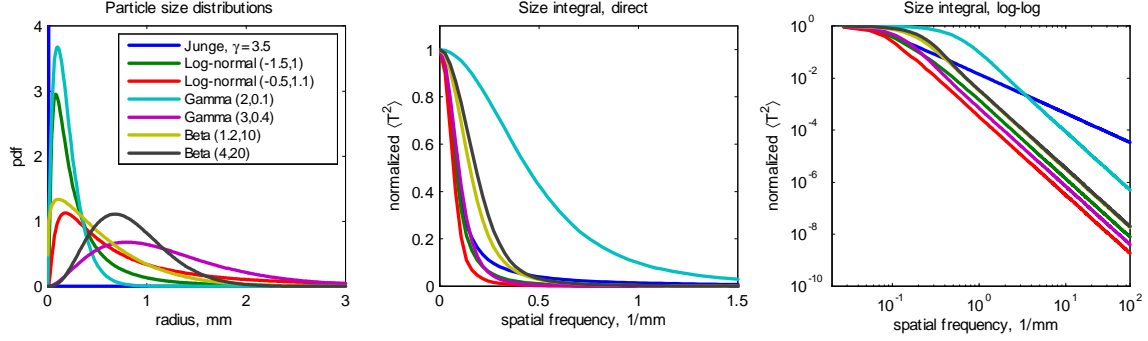


Figure 3-30: A variety of sample particle size distributions using popular models and the associated $\langle T^2 \rangle$ integrals; the integrals have been normalized to $\langle T^2(0,0) \rangle$ for comparison. Values in parentheses after the distribution name indicate the parameters used in Matlab to generate the distribution. The center plot shows that the majority of the energy is within the central low-frequency lobe for nearly all the distributions, despite the variety. The log-log plot indicates that the slope is nearly the same for all the distributions in the mid-to high-frequency ranges, except for the Junge-type distribution and, to an extent, one of the Gamma distributions. The similarity makes distinguishing the distributions or getting more than the mean size difficult from the spectral information alone.

total energy, and averaged together. A plot of the first quadrant of the average power spectrum is shown in 3-31. The spectrum is nearly isotropic¹¹ and decays rapidly with spatial frequency. A fit to the spectrum (averaged over all angles), Figure 3-32, gives an exponent of $a = -2.046$ ($-2.074, 2.019$) with a correlation coefficient of 0.976. Notably, the modulation in the spectral energy from the depth integration also tends to smooth out over enough realizations, and only affects the lower-frequency region. As ρ becomes larger, the sinc modulation of Equation 3.128 goes to a constant and the $\langle T^2 \rangle$ spectrum dominates. One caveat is that nearly all of the holograms were captured in controlled settings in sample tanks, so that bubbles and large particulate matter would have had a greater probability of settling or rising out of the sample volumes under consideration (see, e.g., [380],[379] which found log-normal fits for mixed-phase flows during the settling phase). Thus, a Junge-type distribution consisting of extremely small particles and dissolved matter would have been more likely to occur in these holograms.

The derivations so far have been for planar references. The original premise was that

¹¹The anisotropy is partly a function of the type of detector. Overexposure of pixels in the CCD results in streaking along the columns. Since the CCD was maintained with the same orientation, there was a slight increase in the high spatial frequency content preferentially in the u direction.

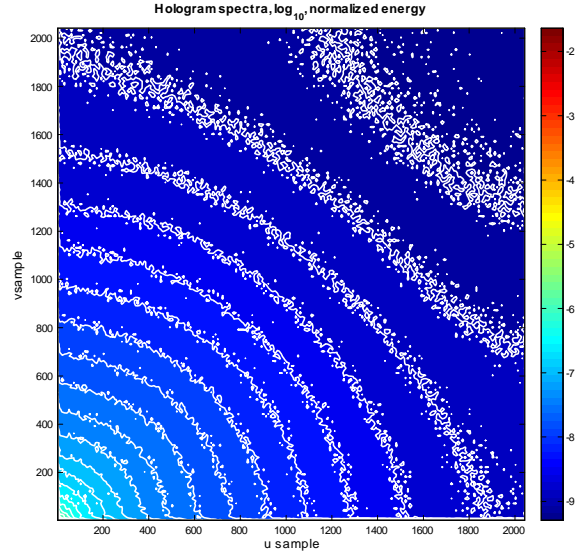


Figure 3-31: First spatial-frequency quadrant of the average spectral energy computed from natural oceanic holograms.

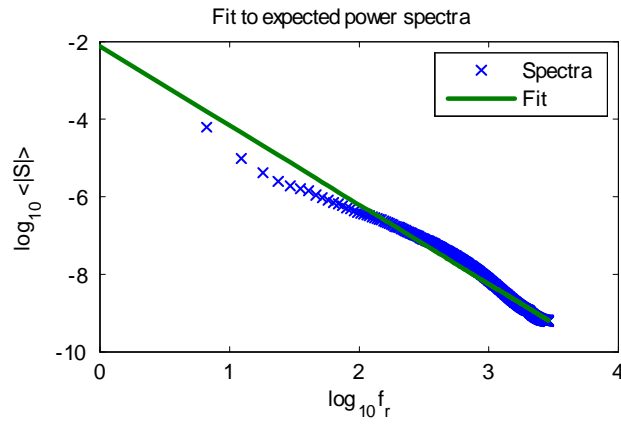


Figure 3-32: Least-squared error fit to the expected power spectrum of Figure 3-31. Here, $f_r = \rho$ and $|S|$ is the spectral energy.

since spherical references differ from planar only in the equivalent reconstruction distance and magnification, the results could be expected to be similar. Further, many different particle size distributions give $\rho^{-\alpha}$ noise, so that the magnification would be expected to re-scale the distribution but not significantly change the noise. Explicitly showing this is more difficult as the depth and shape integrations become linked through their shared z parameterization. The $\langle O^c O^* \rangle$ term, for example, becomes

$$\langle O^c O^* \rangle = \langle s^2 \rangle \int \int p(z, \boldsymbol{\theta}) \exp(i2\pi\lambda\rho^2 z_{eq}(z)) T^*(-u, v; \boldsymbol{\theta}(z)) T(u, v; \boldsymbol{\theta}(z)) d\boldsymbol{\theta} dz, \quad (3.135)$$

where $\boldsymbol{\theta}(z)$ denotes a depth-dependent scaling and z_{eq} is the equivalent reconstruction distance derived from Equation 3.64 using Wigner transform methods. Substituting z_d for the total distance from the spherical source to the detector, $z_{eq}(z) = z_d z / (z_d - z)$ and the scaling factor is $M = z_d / (z_d - z)$ (from substituting $z_d = z_s + z_p$ into Equation 3.63). Assuming a circular particle mask and a uniform distribution over $z = z_a$ to $z = z_b$, the integration becomes

$$\begin{aligned} \langle O^c O^* \rangle &= \langle s^2 \rangle \frac{1}{(z_b - z_a)} \int_{z_a}^{z_b} \exp\left(i2\pi\lambda\rho^2 \frac{z_d z}{(z_d - z)}\right) \\ &\quad \times \int_a p(a) \left[\frac{a\pi z_d}{(z_d - z)} \text{jinc}\left(\frac{a z_d}{(z_d - z)} \rho\right) \right]^2 da dz. \end{aligned} \quad (3.136)$$

Similar equations can be derived for the other components of Ψ_S .

Examples of the normalized spectral energy computed for a spherical reference ($z_d = 388$ mm, $z_a = 35$ mm, $z_b = 240$ mm, $\lambda = 658$ nm) and the same distributions as in Figure 3-30 are plotted in Figure 3-33; solid lines are the normalized spectra from the spherical reference while dotted lines are the same spectra due to a planar reference. Modulation from the depth integral is minimal in the log-log plot, with the dominating term coming from the $\langle T^2 \rangle$ integration. The spectra resulting from spherical and planar references are nearly indistinguishable aside from the offset. The offset itself is due to the fact that $z_{eq} > z_p$, so that the integration for each spatial frequency extends to larger $(z_b - z_a)$ for spherical than for planar.

The results from this section provide strong theoretical support that the PSD of particle

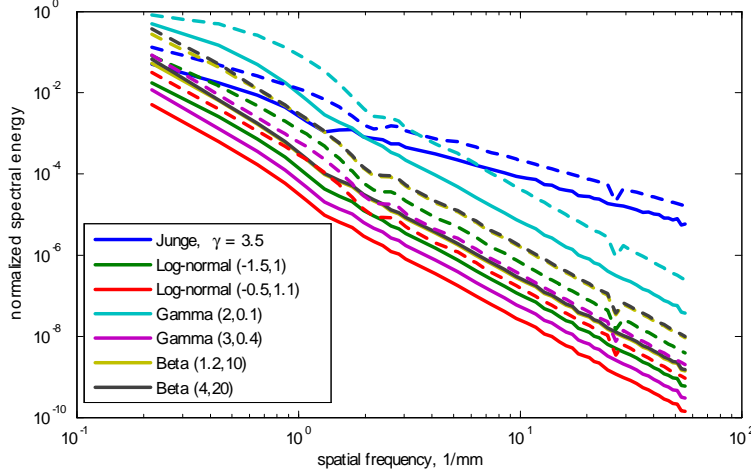


Figure 3-33: Normalized spectral energy curves for various distributions. Numbers after the distribution type in the legend denote the parameters used in Matlab to generate the distribution. Solid lines are the spectra using a spherical reference. Dotted lines are the spectra for the same distributions but using a planar reference. Fluctuations at high frequencies are due to the limited number of z samples used to compute the double integral.

noise follows a ρ^{-a} model independent of the actual distribution. The popular log-normal, gamma, and beta distributions have $a = 3$, while Junge-type distributions have $a = 5 - \gamma$. In particular, a Junge-type distribution with $a = 2$ agrees with the measured PSD of laboratory holograms. Both spherical and planar references result in similar ρ^{-a} models. Unfortunately, since many natural objects follow a similar power distribution, filtering to remove noise may be particularly challenging. In practice, a power filter (Section 3.3.4) with low $p \sim 2 - 4$ visually removes enough particle and discretization noise to make images recognizable, albeit with extremely poor detail.

One way of decreasing particle noise is simply decreasing the sampling volume so that it contains fewer total particles. Writing the expected number of particles in the volume as a function of concentration, C , and cross-sectional area, A , as $N = CAz_{\Delta}$ and substituting into Equation 3.128 gives

$$\Psi_S = 2CAz_{\Delta} \langle s^2 \rangle \langle T^2 \rangle + \frac{2CA \langle s^2 \rangle \langle T^2 \rangle}{\pi \lambda \rho^2} \sin(\pi z_{\Delta} \lambda \rho^2) \cos(2\bar{z} \pi \lambda \rho^2). \quad (3.137)$$

Thus, decreasing the volume affects the oscillation rate of the sinc term but not its decay rate. Attempting to mitigate the effects of the particle noise by changing the sampling

volume changes instead the noise offset.

3.4.1 Image degradation

The actual effect of particle noise on a reconstructed image depends strongly on the spectral content of the object and the imaging goals. For example, man-made objects can have sharp edges and greater high-frequency content, exactly where Ψ_S becomes small. Simulations were computed to illustrate the degradation due to particle fields and to provide an initial step towards linking physical variables with expected image quality.

Particle field simulations were done by propagating an object image through a dense distribution of round particles with a log-normal size distribution. (A log-normal distribution was used to provide insight because of its similarity to beta and gamma distributions, mathematical relationships, and the possibility of limiting errors due to extremely small particles.) The object included a synthetic copepod and groups of horizontal resolution bars with increasing size and distance (similar to a USAF 1951 resolution chart) (see Figure 3-34) and was illuminated with an in-line planar reference. The field was propagated a short distance using the exact kernel (Equation 3.17), then multiplied with a simulated particle mask. The total number of particles per mask was Poisson distributed based on the a particle concentration, the positions were uniformly distributed over the beam, and the radius, R , followed a log-normal distribution given by

$$R = \frac{1}{R\sqrt{2\pi V^2}} \exp\left(-\frac{[\ln R - M]^2}{2V^2}\right). \quad (3.138)$$

The slice model assumes that the particles are distributed independently and that a series of short propagation distances closely approximates a continuous volume. A complete particle field is built by propagating the field after the particle slice by a short distance and repeating the process until the total volume has been simulated. The hologram that would be recorded at the detector is then the intensity of the final field. This direct simulation method has the benefit of including re-scattering and absorption, maintaining some of the higher-order effects.

The hologram was then reconstructed at the known object depth and a number of

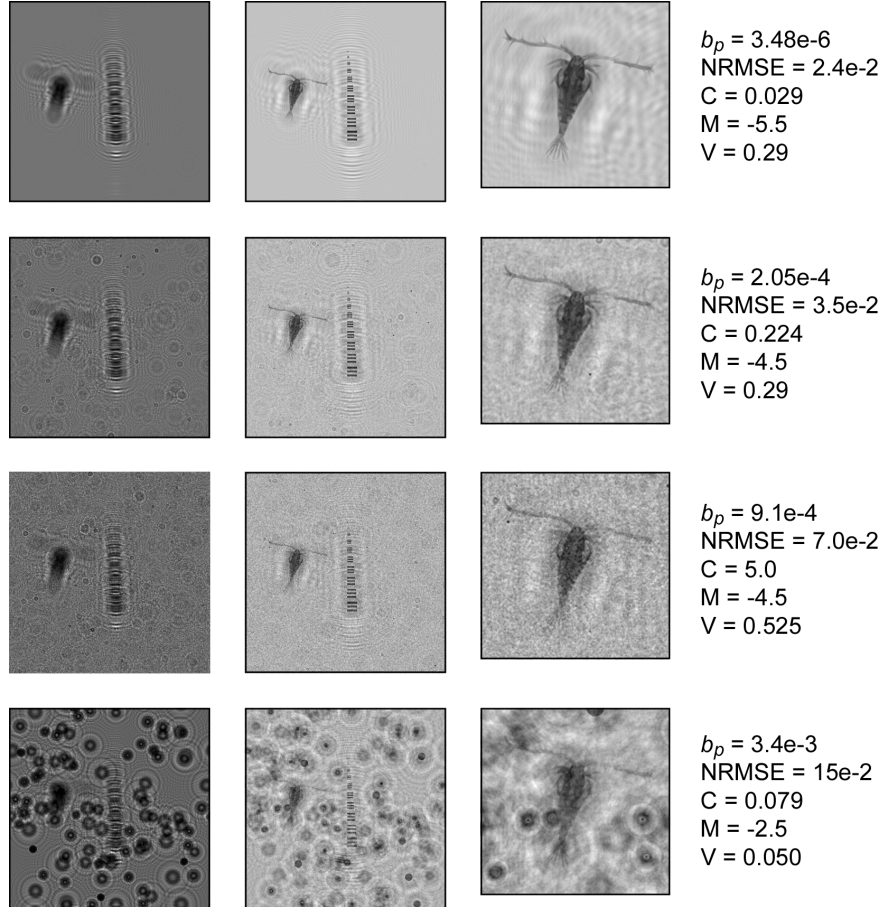


Figure 3-34: Examples of particle simulations with varied distribution parameters M , V , and C . The left column is the hologram, the center column is the reconstruction, and the right column is a larger view of the copepod area. The samples are arranged by increasing b_p and NRMSE from top to bottom.

metrics computed. An area around the copepod was used to estimate the visual image quality using the NMSE (Equation 3.51, or the NRMSE by taking the square root) [116] and a spectral SVD-based metric [333]. (More traditional metrics such as the MSE were also computed for comparison, but discarded as they are dependent on the total energy. Thus, the MSE responds strongly to the total absorption and not the image quality. The NMSE attempts to correct for energy loss.) The average contrast across each set of resolution bars was also computed and stored.

The sample reconstructions in Figure 3-34 illustrate some of the interesting results of the simulation. The resolution bars are clearly visible even when the copepod virtually disappears into the noise. The exact reason why this effect occurs is speculative, but may revolve around higher frequencies in the resolution bar spectra or stronger absorption over the bar area. The resolution, defined as the minimum bar separation that has a contrast above an arbitrary level, remained nearly constant through the simulations and did not have a good fit to the predictor models. Thus, the image quality of the simulated biological object will be reported as the primary metric for the remainder of this section. This also supports the idea of both quantitatively measuring system performance with known engineering methods such as resolution charts and continued testing with real-world objects (Section 3.3.2).

Relating the parameters of the model distribution to measurable values is useful for both practical applications and for understanding the basis of the degradation. Babin et al. follow Spinrad in writing the angle-integrated particle scattering coefficient, $b_p(\lambda)$, as

$$b_p(\lambda) = C \int p(R) Q_b(\lambda, m, R) \pi R^2 dR, \quad (3.139)$$

where C is the number concentration per unit volume, $Q_b(\lambda, m, R)$ is a scattering efficiency derived from Mie theory, and m is the relative refractive index [15]. Walstra approximates Q_b with good accuracy when $m \rightarrow 1$ by

$$Q_b \simeq 2 - \frac{4}{\varrho} \sin \varrho + \frac{4}{\varrho^2} (1 - \cos \varrho), \quad (3.140)$$

where

$$\varrho = \frac{4\pi R(m-1)}{\lambda} = 2q(m-1), \quad (3.141)$$

is a size parameter and $q = 2\pi R/\lambda$ is the particular size parameter commonly used for Mie scattering [408]. A typical value for m is 1.05 for oceanic particles [390], and Waltra's approximation is reasonable. Q_b limits to 2 as $\varrho \rightarrow \infty$, as expected from Mie theory [36], and can be approximated as constant (within 5%) by $R = 40 \mu\text{m}$ for $\lambda = 658 \text{ nm}$. Integrating $Q_b R^2$ in Equation 3.139 smooths out the variations at low R so that the assumption of a constant $Q_b = 2$ results in only a small error for b_p . The estimate becomes

$$b'_p \simeq Q_b \pi C \int p(R) R^2 dR = 2\pi C \langle R^2 \rangle \quad (3.142)$$

$$= 2\pi C \exp(2M + 2V^2), \quad (3.143)$$

assuming $p(R)$ is log-normally distributed or can be approximated by a log-normal. The prime is a reminder that b'_p is an approximation to the scattering coefficient.

The NRMSE is a strong function of b'_p (Figure 3-35). The curve is fit well by

$$\log_{10} NRMSE = c_1 + (c_2 + c_3 \ln b'_p) \left[\frac{1}{2} + \frac{1}{2} \tanh(c_4 \ln b'_p - c_5) \right], \quad (3.144)$$

with coefficients given in Table 3.4; the fit has an adjusted $r^2 = 0.9905$. The c_1 constant is the asymptotic limit due to the twin image term. For extremely low b'_p , $\ln(b'_p) \lesssim -9$ (around $b'_p = 1 \times 10^{-4}$) the image is degraded primary by additive noise from nearby defocused particle images. As b'_p increases, particle scattering begins to cause more significant information loss, modeled as the linear term. The tanh function in the square brackets is an approximation to the Heaviside function and acts to switch smoothly between the asymptotic value and the loss function.

Future work with particles would connect the initial theory developed here to the exact scattering and absorption coefficients to anticipate the oceanic areas where holography can be used given a desired imaging resolution or quality. The Office of Naval Research maintains an online database of oceanic optical properties that links worldwide locations with known measurements [265] that should in this prediction. The actual scattering is

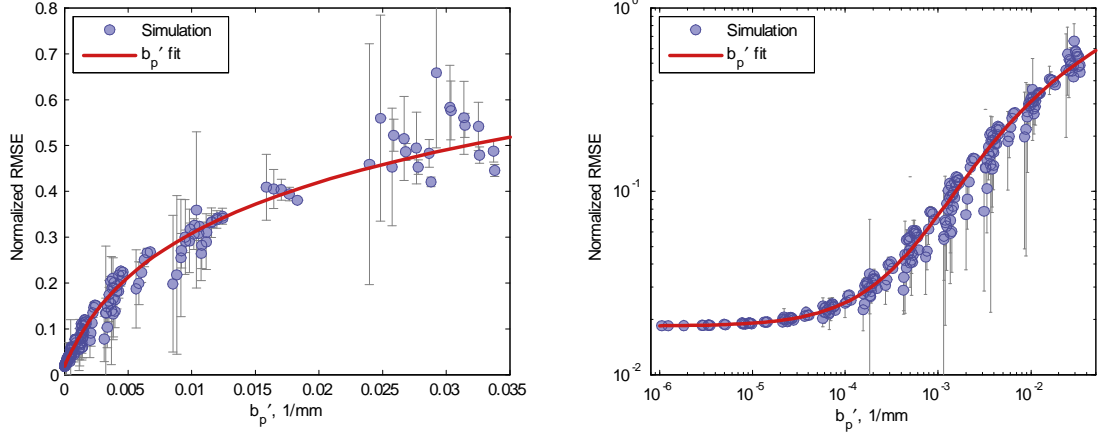


Figure 3-35: Plots of the normalized RMSE (square root of NMSE) as a function of b'_p . Error bars are plotted for the simulation values ($\pm 1\sigma$) and are calculated based on repeated trials with the same distribution parameters.

Coefficient	Value	Interpretation
c_1	-1.735	twin image asymptote
c_2	1.87	loss intercept
c_3	0.105	loss rate
c_4	0.416	switching rate
c_5	-2.93	switching center point

Table 3.4: NRMSE prediction coefficients

a non-trivial combination of biological and physical phenomenon [256],[330],[390], making this a potentially interesting and involved project in its own right.

This chapter has focused on the practical development of a prototype unit for capturing *in situ* holograms of plankton and other microscale objects and understanding how engineering decisions affect the imaging. Theory was developed to understand how motion would affect the holographic images and provide heuristics for limiting the motion. A Wigner analysis was used to describe the fundamental capabilities of spherical reference holography, including the resolution and bandwidth limits, the sampling volume, and depth of field. Once the prototype had been tested and constructed, methods of quickly reconstructing images from Bayer-filtered cameras were created. The final section discussed theory regarding particle noise, a challenge for any real-world holographic system.

Of course, capturing images is only half of the challenge. The next chapter discusses a number of new methods for extracting the information and images hidden within the prototype holograms. This has traditionally been one of the greatest limiting factors for the use of digital holography. Providing faster, more reliable, and better understood methods goes a long way towards elevating digital holography as a widespread tool.

Chapter 4

Extracting Information and Images from Holograms

Photography was developed over a long period starting in the 1820's, becoming more popular and useful as the tools improved over the 19th century [303],[422],[375]. Digital holography faces a similar expansion: the basic technique has developed over the past several years but faces limited use as researchers struggle with algorithms to reconstruct images and extract information from holograms in a timely and reliable way – with enough simplicity that the tool can be adopted by users who do not have specialized knowledge in optics, signal processing, and computer science. The first hurdle to widespread use has been in computational power, quickly increasing from the early 1980's when it took minutes to reconstruct digital holograms and practical sizes were limited to 256 pixels per side or smaller to the latter part of the 2000's when holograms of 4096 pixels per side can be reconstructed multiple times per second using off-the-shelf hardware¹ (Section 4.2). The second major hurdle has been in the algorithms used to extract information from the holographic images, which vary considerably based on the application and method of recording the hologram. Certainly a digital hologram has more information available than a standard image, such as having access to the real and imaginary components, the ability to arbitrarily refocus an optical field, and having the object depth encoded into the spectrum. Making use of

¹Despite the computational capabilities, most digital holography users are still content in 2010 to limit themselves to 1 – 2 MPx holograms and slower processing methods on personal computers.

this additional information requires a good understanding of the physics and computational methods. It is also worth noting that some traditional image processing techniques can be used with holographic images, but others face challenges. For example, binary thresholding leads to poor object-background separation due to the twin image of in-line holography. The end result is that algorithm development has been somewhat limited.

The various attempts at applying DH to marine studies are no different than laboratory holography and have also consistently struggled with software. Owen and Zozulya, Katz et al., and Kruezer et al. have reported that they have tools for working with holograms. In talking with Malkiel and Kruezer, the tools are aimed at human-guided reconstructions and working with individual holograms, so that automatically retrieving data from the images (especially in a robust and time-efficient manner) is a large part of what halted some of the various historical underwater DH projects (Table 3.1). The early devices and tools also took a greater understanding of holography and the processing methods to effectively gather data in the first place, limiting adoption by a wider audience.

This chapter examines a number of algorithms, techniques, and methods which have been created to work with in-line holograms, extending the toolbox available to holographers. The first section aims at locating objects within the 3D volume using focus metrics applied to reconstructed images or through fast position estimators. The second section talks about how a commercial graphics card can be used to significantly increase the computational rates and which algorithms are more appropriate for implementation given the restrictions imposed by the hardware. The final section presents a case study where *in situ* digital holography was used to measure oil droplets, utilizing a focus metric that was more readily applied to fast hardware reconstructions for determining candidate droplet positions with extremely large datasets.

4.1 Detection, position estimation, and focus determination

One of the motivating factors for using digital holograms in science is that large volumes of spatial information can be recorded in a single shot, then reconstructed at a later time to examine the objects within the sampled volume. The reconstruction requires the 3D

position of the object of interest, the lateral position usually within a few pixels and the depth within approximately the depth of focus around the object. Some applications and objects may be more lenient than others, so that the estimation need not be perfect. Other applications which require highly focused estimates, such as exacting size determination, benefit from further position refinement. The question that this section addresses is how those position estimates and refinements can be done in the first place.

Object detection, position estimation, and focus determination are strongly linked. Intuitively, an object is located at the position where it is best in focus. Similarly, an area where several edges are in focus denotes the presence of an object. Detection is the process of deciding that there is enough evidence of an object or an edge to denote an object. The distinctions in this section are somewhat arbitrary, and Section 4.1.3 intentionally merges detection and focus metrics to create a novel method of finding edges in real-world holograms.

A common technique for focus determination follows the idea of an auto-focus in a camera. Multiple holographic reconstructions are computed and a focus metric is applied to each slice. The metric highlights focused objects, providing simultaneous detection and lateral positioning. The depth of the strongest focus response determines the object's depth. The challenge with these combined detection-position methods is selecting good metrics which have selectivity for the objects and measurements of interest. For example, high position sensitivity, necessary for measurements of tracer particle velocity in fluid flow experiments, requires a metric which has a narrow peak in the depth direction. Different focus metrics are discussed in the first part of this section, including new techniques which give better depth and lateral estimates than current methods.

Reconstructing the entire volume can be extremely time consuming, especially when the objects are sparsely distributed. Methods which separate the detection, lateral position estimation, and depth estimation can be used to quickly hone in on objects, leading to much faster retrieval of images from the source holograms. Focus metrics can then be applied to refine the estimates. The latter part of this section discusses these fast estimation techniques.

4.1.1 Traditional focusing methods

Image processing has a long history with focus detection, so it is not surprising that the first methods used by holographers followed those of the imaging community. The traditional method works similarly to the auto-focus in a camera: the focus is scanned over some range, and a metric is computed which relates directly to how well the image appears to be in focus at each focal position [206].

The most basic measure of focus is to examine the edge sharpness. This is especially useful for in-line holography where the object is often defined only by its edges due to high absorption, high scattering angles, or multiple scattering within the object boundary. The edge sharpness is measured directly by the magnitude of the gradient. Filters such as the Prewitt or Sobel (see Figure 4-1) are used to compute the gradient in the x and y directions, with the edges occurring along the largest gradient magnitude. Simple thresholding provides a straight-forward decision about whether an edge exists at a specific pixel location; more intelligent thresholding, such as Canny's hysteresis thresholding, can help link together strong edge detections with weaker edge information [54],[368].

Edge filters such as Prewitt's and Sobel's are well-known to be sensitive to both noise and the gradient direction [368],[323]. A steerable gradient filter can be used instead to calculate the edge strength, defined as

$$S = \frac{-2x}{\sigma^2 n^2} \exp\left(-\frac{x^2 + y^2}{\sigma^2}\right), \quad (4.1)$$

with σ acting as a filter width parameter, n is the number of pixels and serves to act as a simple normalizing factor, and x and y are the distances of the pixels from the center of the filter (Figure 4-1). Steerable gradient filters have two significant benefits: the obvious one is that the Gaussian like smoothing reduces noise sensitivity (at the cost of decreasing position sensitivity). The other benefit is that the edge gradient calculated at an angle θ , S_θ , is a linear combination of the filter responses in the x and y directions, S_x and S_y respectively, as

$$S_\theta = S_x \cos \theta + S_y \sin \theta, \quad (4.2)$$

so that only two filtering operations are necessary to compute the gradient in any arbitrary

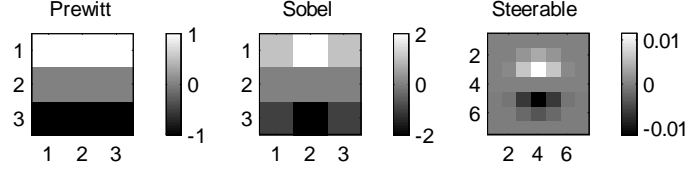


Figure 4-1: Comparison of Prewitt, Sobel, and steerable gradient edge filters. A σ of 1.5 was used for the steerable edge filter. All three filters are sensitive to horizontal lines.

direction (hence the name “steerable”) [126]. Further, the maximum gradient occurs at $dS_\theta/d\theta = 0$, or $\theta = \arctan(S_y/S_x)$, and substituting this value into S_θ gives the maximum gradient magnitude as $S_M = [S_x^2 + S_y^2]^{1/2}$. A steerable gradient filter is preferred over traditional gradient filters for its noise characteristics and lack of directionality artifacts.

A Laplacian filter, the second derivative, is another common metric for edges: zero crossings indicate the extrema of the gradients. The crossings can be determined by computing the local minimum and maximum around pixels with low Laplacian magnitudes and verifying that a zero crossing is possible within the nearby area. Local minimum and maximum can be found quickly with grayscale morphological operations [402]. The Laplacian is especially sensitive to noise as a second derivative filter, and is commonly combined with a Gaussian filter for local smoothing [368].

Domínguez-Caballero deviated from the norm and searched instead for areas with high Laplacian magnitudes for his focus metric [95]. The reason why this works is that the second derivative has a high value *near* the edge. Consider a true edge as a Heaviside step function between a dark object (low value) and bright background (high value). Approximating a 1D Heaviside using a continuously differentiable function, the hyperbolic tangent [12],

$$H_a(x) = \frac{1}{2} + \frac{1}{2} \tanh(kx) = \frac{1}{1 + e^{-2kx}}, \quad (4.3)$$

with higher k indicating a sharper transition at the edge, the gradient is

$$\frac{dH_a}{dx} = \frac{2ke^{-2kx}}{(1 + e^{-2kx})^2}. \quad (4.4)$$

The second derivative (the Laplacian), $L(x)$, is

$$L(x) = \frac{d^2 H_a}{dx^2} = \frac{8k^2 e^{-4kx}}{(1 + e^{-2kx})^3} - \frac{4k^2 e^{-2kx}}{(1 + e^{-2kx})^2}, \quad (4.5)$$

$$= \frac{4k^2 e^{-2kx}}{(1 + e^{-2kx})^2} \left[\frac{2e^{-2kx}}{1 + e^{-2kx}} - 1 \right]. \quad (4.6)$$

The term in the square bracket goes to zero at $x = 0$, but quickly shoots back to its limits of $[-1, +1]$. The extrema of $L(x)$ occur at

$$x_E = -\frac{\ln(2 \pm \sqrt{3})}{2k}, \quad (4.7)$$

so that as the edge becomes sharper and k increases, the extrema move closer to the true edge. The magnitude of $L(x)$ at the extrema is

$$|L(x_E)| = \frac{4k^2 (2 + \sqrt{3}) (1 + \sqrt{3})}{(3 + \sqrt{3})^3} \simeq 0.3849k^2, \quad (4.8)$$

which increases quadratically as the edge sharpness k increases. Compare this against the gradient: while the gradient extrema is centered above the edge, the gradient magnitude goes linearly as $k/2$. The quadratic factor gives the maximum Laplacian magnitude metric greater depth discriminability than the maximum gradient magnitude as the edge comes into focus. It should be noted that Tachiki also uses the magnitude of the Laplacian as a focal metric, integrated over a local neighborhood [369] – with the spatial integration removing a degree of the discriminative benefit of the Laplacian.

The local variance or standard deviation serves as a proxy for a range measurement. It has seen use in the past for shape-from-focus techniques using *ad hoc* arguments for its effectiveness [235],[369],[248]. The local variance is small when the intensity is nearly uniform, and grows as the intensity range increases. The local variance has a benefit of reaching a fixed maximum, which occurs when half the pixels in the local neighborhood used for computing the statistic are dark while the others are bright. The action of using the local variance can be seen by again using H_a to approximate an edge coming into focus. The local variance in a neighborhood of size S around an edge is

$$Var_S \{H_a\} = \frac{1}{S} \int_{-S/2}^{S/2} \left(H_a(x) - \frac{1}{2} \right)^2 dx \quad (4.9)$$

$$= \frac{-2e^{kS} + 2kS + kSe^{kS} + kSe^{-kS} + 2e^{-kS}}{4kS(1 + e^{-kS})(1 + e^{kS})}$$

$$= \frac{1}{4} - \frac{\sinh(kS)}{2kS[1 + \cosh(kS)]}. \quad (4.10)$$

Taking the Taylor expansion,

$$Var_S \{H_a\} \approx \frac{1}{48}k^2S^2 - \frac{1}{480}k^4S^4 + O(k^6), \quad (4.11)$$

shows that the local variance increases approximately quadratically with k as the edge initially comes into focus. Similarly, the Taylor expansion of the root of the local variance, the local standard deviation, is

$$\sigma_S \{H_a\} \approx \frac{\sqrt{3}}{12}kS - \frac{\sqrt{3}}{240}k^3S^3 + O(k^5), \quad (4.12)$$

which, as expected, shows a linear rate for small k .

The key for local variance is the shape of the response as the edge comes into focus, plotted in Figure 4-2. After the initial growth rate of k^2 , the $\sinh(kS)$ and $\cosh(kS)$ terms of Equation 4.9 nearly cancel out, so that Var_S goes like

$$Var_S \simeq \frac{1}{4} - \frac{1}{2kS} \quad (4.13)$$

for large k . This means that the local variance has a good sensitivity to the initial formation of the edge, then quickly stabilizes at a known maximum. The result is a metric which is reliably near a known maximum value over a range of depths, then quickly disappears: a depth-selective window, in a sense.

Also popular for holography is the use of minimum intensity as a metric [380],[11],[103]. Objects which absorb or scatter at high angles appear dark in reconstructions. As the objects go out of focus, the energy is spread into the nearby area and the intensity tends towards the local mean. The minimum intensity works well for particles, but is not guaran-

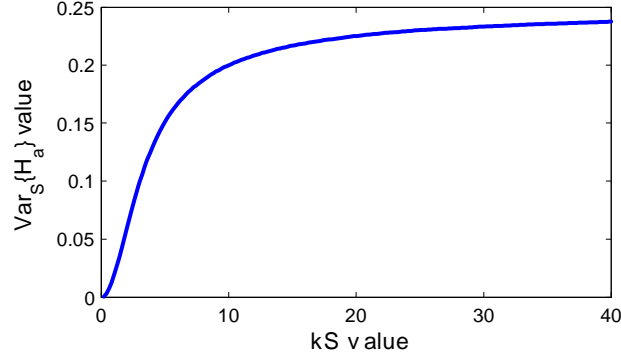


Figure 4-2: Sample plot of the local variance, Equation 4.9, as the edge comes into focus.

teed to work for larger objects. Figure 4-3 shows the intensity cross-section through a point object as it defocuses, with a distinct minima at the object. Note that a darker cone extends out beyond the point object, so that points near to the object have minimum intensities at defocus greater than zero, and thresholding is necessary to limit the focus metric to the correct region. Figure 4-4 shows the intensity cross-section through a larger *rect()* object with partial transmission. Pixels near the rectangle edge have minimum intensities at defocus significantly different from zero, marked at lines B, C, and D – and more importantly, a lower intensity than the object at focus. The Wigner transforms (see Equation 3.54 and related discussion) for each of the defocus lines are plotted in Figure 4-5. The Wigners have alternating bands of positive (red) and negative (blue) values, following $1/|x|$ -shaped bands. As the rectangular signal defocuses, the Wigner shears slightly, causing significant portions of the negative bands to shift into single pixels, with a reduced contribution from the corresponding positive band. The result is that the intensity at the particular pixel drops, leading to errors in the edge and depth measurement. The point object is narrow enough that the positive and negative bands are nearly vertical, and thus the effect is not seen within the reconstruction pixel.

Two groups of mention have attempted to find focus metrics specific to holograms. Meng et al. have proposed additional sets of focus metrics which make use of the complex field present in a hologram [89],[274]. Each of these methods is developed for use with particles. Objects larger than a few pixels have unreliable responses to these focus metrics, and they are not discussed further here. Liebling et al. has proposed an alternate reconstruction

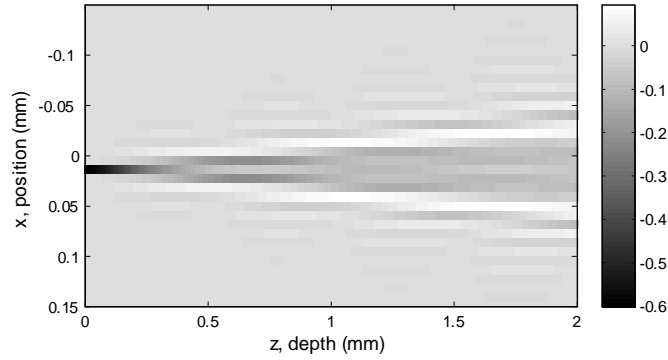


Figure 4-3: Intensity cross-section through a point object as a function of defocus in the z direction. The scale is \log_{10} of the intensity.

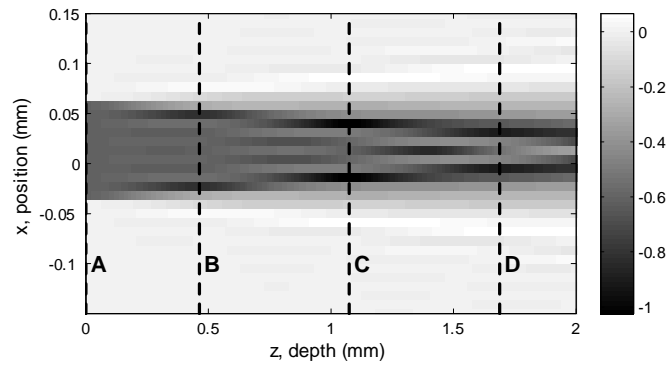


Figure 4-4: Intensity cross-section through a rectangular object as a function of defocus in the z direction. The scale is \log_{10} of the intensity for improved visibility of the minima. Vertical lines correspond to minima at pixels close to the object edge, each with intensities lower than the actual edge pixels. The lines match the plots of Figure 4-5.

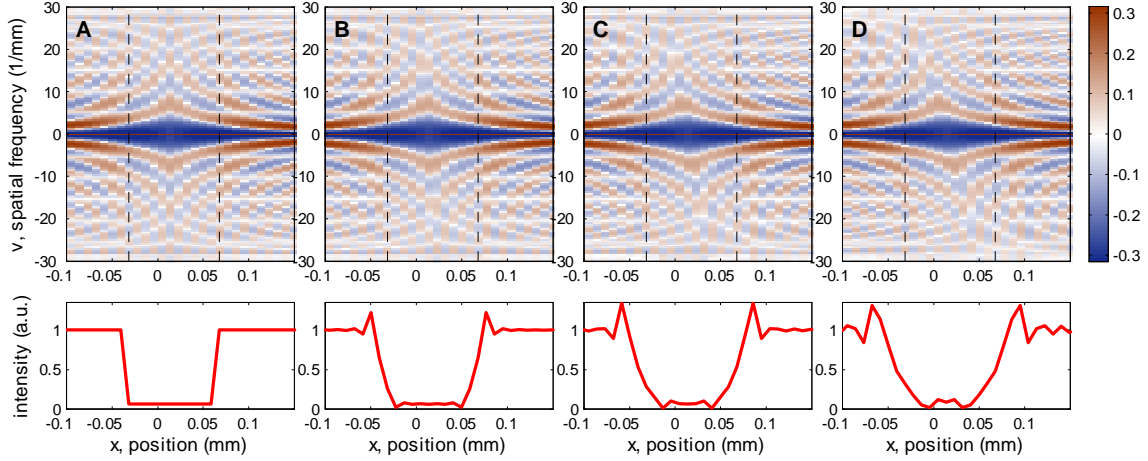


Figure 4-5: Wigner transforms and intensity signals through the profiles denoted in Figure 4-4. The Wigner transform is zoomed to the central portion and the magnitude scaled by a signed square root for visibility. Dashed lines in the top plots denote the edges of the original *rect()* signal. The observed intensity is plotted below each Wigner.

technique for holograms that involves decomposing a hologram into a wavelet basis which has been Fresnel transformed (termed “Fresnelets”), with a different Fresnel coefficient for each reconstruction depth [221]. They create a focus metric by computing the energy² of the strongest wavelet coefficients (to an arbitrary number or percentage) at each test depth, with the claim that in-focus objects require few Fresnelets for representation [222]. The resulting focus metric measures the energy in the focused regions, making it related to a localized minimum intensity type of metric. The depths with “sparsest” coefficients can be reconstructed to estimate the objects and positions, or the Fresnelet coefficients can be used to estimate the lateral position. There are two critical caveats for the Fresnelet method. The first is that the Fresnel transform is additive in the complex domain, so that the hologram’s complex field must be measured and used for the Fresnelet decompositions – and is thus not applicable to single-shot holograms. The computational cost can also be surprisingly limiting, on a similar order as FFT-based reconstruction. The second caveat is that Liebling assumes the object image can be represented by a sparse wavelet basis, not always the case for real-world objects.

²It may be interesting to note that Liebling et al. use the ℓ_2 energy of the strongest wavelets to estimate the ℓ_0 sparsity (albeit with a limited number of coefficients). This is different than compressive sensing, a hot topic in both computer vision and holography, which uses an ℓ_1 norm.

The traditional metrics discussed in this section can be applied either by scanning through the depth and recording the z where each pixel maximizes its metric (Figure 4-6), or by finding local neighborhoods in z where each pixel (or small group of pixels) maximizes its focal metric. The latter method has the downside that the number of maxima is not known ahead of time, so that each detection grows the memory requirements. This can be problematic when processing on a dedicated board (for example, a GPU or other DSP board) where memory is limited and the operations are significantly faster if there are no branches (see Section 4.2). The first method of searching for pixel-wise maxima across all depths has a predictable and fixed memory overhead with the same operations applied across all pixels and is thus ideal for DSP boards. This process can be thought of as a maxima projection in the depth direction.

A pixel-wise maxima of the focus metrics assumes that each pixel responds primarily to a single object – in other words, that the sampling volume has a sparse enough population so that only one object is found at each pixel location in the projection. Given one object, the probability p_o that another object overlaps its projection is approximately

$$p_o(r_1, r_2) = \frac{\pi(r_1 + r_2)^2}{D^2}, \quad (4.14)$$

where D is the edge length of the projection area and the objects are approximated as circles with radii r_1 and r_2 uniformly distributed over the $D \times D$ square. This also assumes that r_1 and r_2 are small compared to D so that edge effects are negligible and that the positions of the two objects are statistically independent. The probability that the given object does not overlap *any* of the $m-1$ other objects in the scene is $(1 - p_o)^{m-1}$. Similarly, the probability that *none* of the m objects overlap is

$$p_m = (1 - p_o)^{\binom{m}{2}}, \quad (4.15)$$

with $\binom{m}{2}$ denoting the number of pair-wise combinations to consider; this includes an implicit assumption that the possibility of overlap between three or more objects is negligibly small.

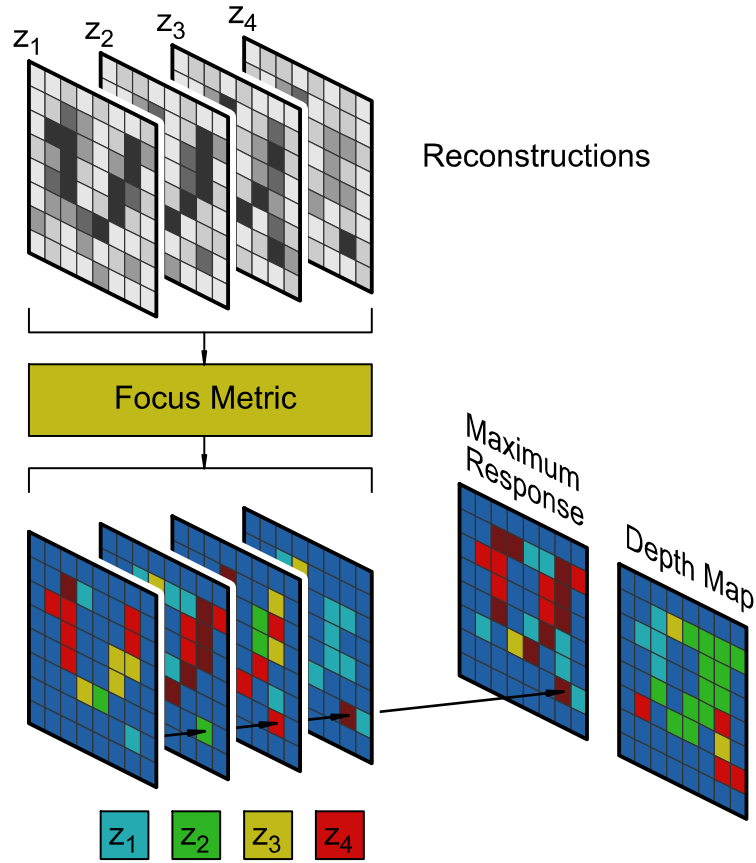


Figure 4-6: Projection method for determining where each pixel maximizes its focus. The focus metric is applied to each reconstruction and compared against the previously best-known maximum at each pixel. If the new reconstruction slice has a better focus, the value of the response and the depth are recorded for the pixel. A map of the maximum response and a depth map are created through the process.

If a uniform distribution between r_{\min} and r_{\max} is assumed for the radii, p_o becomes

$$\begin{aligned} p_o &= \frac{1}{D^2} \int_{r_{\min}}^{r_{\max}} \int_{r_{\min}}^{r_{\max}} \left(\frac{1}{r_{\max} - r_{\min}} \right)^2 \pi (r_1 + r_2)^2 dr_1 dr_2, \\ &= \frac{\pi}{6D^2} \left(\frac{1}{r_{\max} - r_{\min}} \right)^2 (7r_{\max}^2 + 7r_{\min}^2 + 10r_{\max}r_{\min}), \end{aligned} \quad (4.16)$$

found by computing the expected overlap probability. More realistic distributions for the radius, such as log-normals, do not have a good closed-form solution for their integrations and p_o must be computed numerically.

Spherical reference holography includes an additional magnification factor as a function of the depth, $M(z)$ (see Equation 3.63 and related discussion). The observed object size at the projection plane is then $r_{obs} = r_1 M(z)$. The distribution of $p(r_{obs})$ becomes a derived distribution on $p(r_1)$ and $p(M(z))$. The $p(M(z))$ function itself is challenging to evaluate directly because it depends on assumptions for the distribution of objects over the depth. Assuming a uniform distribution of samples over a volume between $z = z_a$ and $z = z_b$ with a spherical reference source at a distance $z = z_d$ from the detector, the distribution for $p(M(z)) = p(M)$ is given by

$$p(M) = \frac{1}{M^4} \frac{3M_b^3 M_a^3}{(M_b^3 - M_a^3)}, \quad (4.17)$$

where $M_a = z_d / (z_d - z_a)$ is the magnification factor at z_a and M_b is similarly the magnification at z_b ; the cubed terms come from integrating over a pyramidal sample volume. The effect of magnification becomes less pronounced for smooth size distributions as the derived distribution for r_{obs} differs less from the distribution on r .

Plots of p_o and p_m for an example case of $m = 20$ objects are shown in Figure 4-7. Log-normal distributions for the object sizes, a value of $D = 36.9$ mm for the detector square (4096 pixels with a 9 μ m pitch per side), and volume parameters matching the prototype unit were used for the calculations. The log-normal parameters were selected to be similar to observed phytoplankton and zooplankton distributions [350],[186],[133],[306]. The probability of there being no overlap between any of the objects is relatively high, 75-90%, even with large objects, so that the overall risk of overlap in a projected focus metric

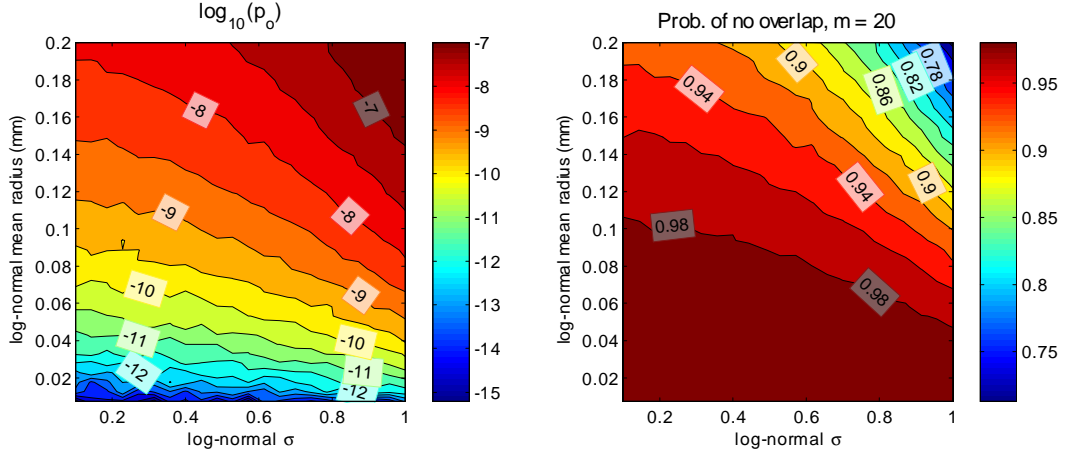


Figure 4-7: Probabilities of object overlap in a projected (pixel-wise) focus metric with spherical reference geometry and log-normal object size distributions. Left: $\log_{10}(p_o)$ as a function of the distribution parameters. Right: p_m , the probability that *none* of $m = 20$ objects will overlap in the projection.

is small.

The estimates for probability of overlap do not include the severity, merely that the objects intersect at some level within the projection image. A portion of the perimeter needs to be visible from each object for it to be detected in the projection using edge focal metrics. Tian et al. demonstrate a way of estimating which pixels in an overlap belong to different objects using a Gaussian mixture model if the objects have a depth separation greater than the uncertainty of the pixel-wise depth estimate [380].

4.1.2 Combined focus metrics

Individual metrics such as the image gradient or the minimum intensity have the benefit of intuitive derivations and close relationships to physical phenomenon. Improvements are possible by combining metrics together, at the loss of direct physical intuition. Combinations based on products of individual metrics benefit from ease as compared to linear combinations, as weights for each metric do not need to be determined.

One of the simplest combinations that will be used later in Section 4.3 is the product of the edge gradient and local intensity. Requiring a large gradient magnitude at the same time the image intensity is small emphasizes only those gradients which belong to dark

objects. The locality condition is necessary since gradients occur at the transition between light and dark regions, and the pixel with the largest gradient would thus correspond to a mid-level intensity value. A morphological filter is used to compute the local minimum as the smallest value in a 3×3 , 5×5 , or similar neighborhood.

The intensity locality condition can be combined with the gradient filtering by using a filter with a larger spatial width, such as the steerable gradient. The filter response is then spread out over a larger area, similar to the area on which the morphological filter would have operated. The resulting focal metric, abbreviated as *SIM* (*Steerable-Intensity-Maximum*), is

$$SIM(x, y) = S_M(x, y) [I_{\max} - I(x, y)], \quad (4.18)$$

where I is the intensity at pixel (x, y) of the reconstructed slice, S_M is the magnitude of the image filtered by the steerable gradient, and I_{\max} is a constant intensity larger than any $I(x, y)$ value. I_{\max} is chosen to provide a large $(I_{\max} - I)$ value when $I(x, y)$ is small. This is preferable for GPU implementations, as direct filtering is straight-forward and faster than the naïve implementation of morphological filters. Restricting the locality also allows for fast, pixel-wise operations.

Other combinations are, of course, feasible. The individual metrics of the previous section, summarized in Table 4.1, were combined using all possible groupings for a total of 57 new metrics and tested on sample holograms. The individual metrics are shown in Figure 4-8 for a copepod in a particle field; the copepod is at a distance of 40.6 mm. The gradient and local Laplacian both have more cloud-like noise around the copepod from twin image and defocusing effects. Defocusing leads to particularly strong gradients a few millimeters from the actual focus, Figure 4-9, leading to depth estimation errors from these edge-based metrics.

Measuring the performance for the metric combinations is highly dependent on the final application goals. Extracting particles for holographic particle imaging velocimetry requires extremely high depth accuracy and few false positives [160],[273]. Obtaining images of animals requires few false negatives, while false positives can be removed using additional steps in the image recognition engine. A simulated hologram with multiple objects at known depths was used to examine the different metric combinations, recording the expected depth

Index	Metric	Notes
1	$I_{\max} - I(x, y)$	Intensity: dark objects have high value; phase objects have little response; can help group spatially separated edge responses
2	$ \nabla I(x, y) $	Gradient magnitude: proportional to k
3	$ \nabla I(x, y) ^2$	Squared gradient: proportional to k^2 for sharper response
4	$ \nabla^2 I(x, y) $	Laplacian magnitude: proportional to k^2 , slightly offset from edge
5	$\sum_S \nabla^2 I $	Local Laplacian: removes some of the spatial dependence of the Laplacian; follows [369]
6	$\sum_S (I - \mu_I)^2$	Local variance: acts like a window in depth

Table 4.1: Individual metrics used in combination

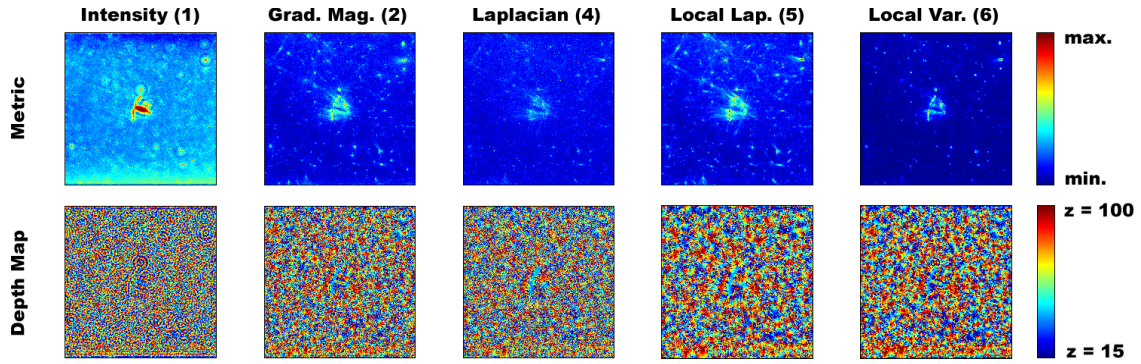


Figure 4-8: Individual focus metrics applied to a test hologram. (Squared gradient is not shown.) The maximum value of the metric at each pixel is shown in the top row, while a depth map of where the metric maximized is plotted in the lower row.

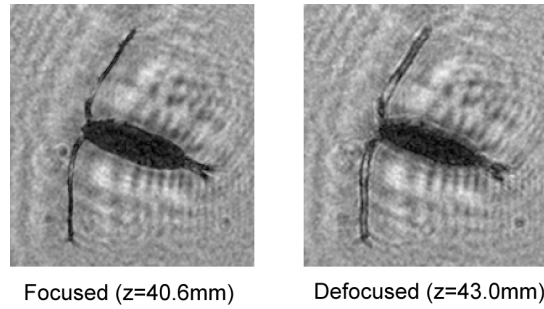


Figure 4-9: Reconstruction magnitudes for a copepod in focus (left) and slightly out-of-focus (right). Diffraction from the thin antennae leads to strong gradients in the defocused image.

error, the number of correct detections, and the total number of detections.

The method for detecting objects from the maximum focus metric image can make a significant difference. A number of approaches were tried here. First, the focus metric, FM, is either used directly or scaled by $FM^{1/m}$, where m is the number of constituent individual metrics. Next, the local mean, μ_{S+} , and standard deviation, σ_{S+} , of the adjusted FM are calculated with large spatial bandwidths and a z -score, $FM_z = (FM - \mu_{S+}) / \sigma_{S+}$, is used to normalize against slow variations in the hologram statistics. The FM_z is then thresholded at two levels: a low level to find objects, and a high level to find reliable edge pixels for the object. A hysteresis-like thresholding is done with the low-threshold objects and high-threshold edges, grouping together all the edge detections which correspond to the same object and eliminating objects which do not have strong edge detections. The collection of edges for each object are then used to estimate depth by using one of three methods: selecting the depth of the pixel with the maximum FM_z [95], a weighted average using all pixels and weights determined by FM_z , and a weighted average with the highest 20% of FM_z values.

Results vary significantly depending on the hologram power signal-to-noise ratio (SNR) and the threshold levels. Synthetic holograms with no additional noise had nearly equal depth estimation performance independent of threshold level or scaling, with performance varying depending on the metric and how the depth was estimated. Scaling the FM leads to lower numbers of true positives, while an un-scaled FM gives more false positives. As the SNR decreases, lower threshold values and a scaled FM lead to better depth results for all three estimation methods. Using four or more individual metrics in the combined metric tends to lead to higher false negatives. Three or fewer individual metrics result in marginally better depth estimation (see, e.g., Figure 4-10, particularly the blue \times and green \square symbols for un-scaled FM). For most metrics and SNRs tested, the lowest expected depth error comes from using the entire set of detected edge pixels to compute a weighted average (see, e.g., the red \bigcirc symbol of Figure 4-10), with a number of combinations performing near the average. Combinations which performed particularly poorly for depth estimation at low SNR included minimum intensity and one of the two Laplacian metrics, while poor depth estimation at high SNR came from metrics that included gradients and local variance.

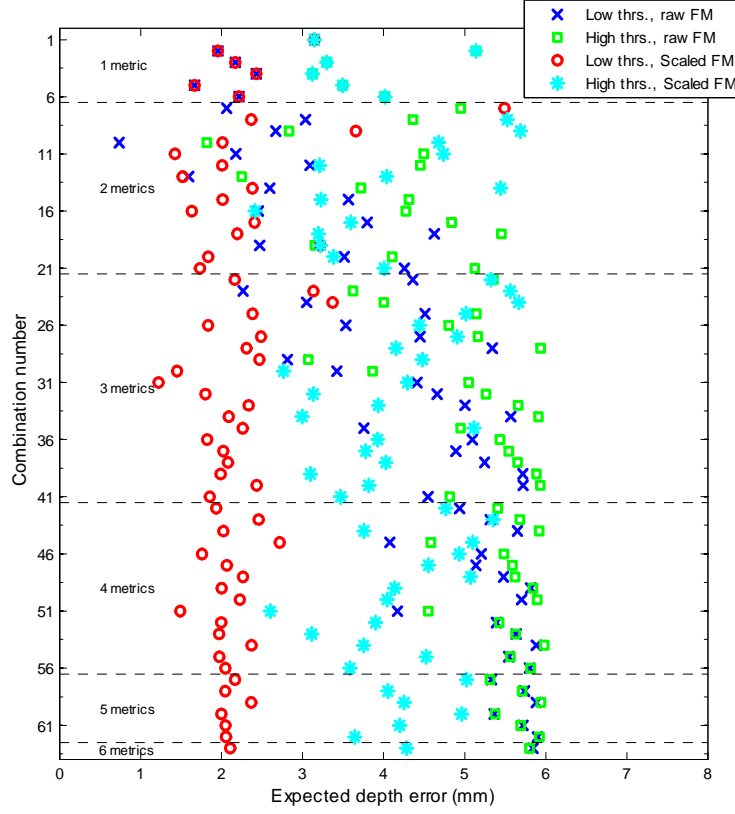


Figure 4-10: Expected depth error for a synthetic hologram with $SNR = 1.0$ and five objects distributed between 45 and 65 mm for each of the different combinations of metrics. The depth is estimated by a weighted average of the edge detections for each of the objects. Dashed lines denote the regions for which specific numbers of metrics were used to form the combination. The data from the red circles had better expected errors for this SNR, but also resulted in missed detections.

The general conclusion is that the individual edge metrics vary, but using a set of lower thresholds and a higher percentage of the detection pixels to estimate depths are both beneficial. Any number of appropriate metrics and thresholds are available for a given task and should be tested for the specific goal and expected SNR.

It is worth repeating that the depth error is dependent on the detection problem of selecting which pixels to trust as edges. Section 4.3 shows improved results for real-world holograms and smoother edges. Segmentation methods may also be useful in future work to improve region and edge predictions.

4.1.3 Metrics using machine learning

A second form of combining metrics comes from the field of machine learning. The guiding principle is that human intuition only goes so far, and tends to focus on finding specific patterns. Computers have the ability to forego preconceived patterns and search for discriminative features amongst a slew of data which might mean little for a human. The role of the user is to provide training samples and determine which algorithms are adequately selecting for the desired results.

A database of training samples was compiled for the task of detecting in-focus edges. A set of 1121 in-focus edge pixels (with corresponding reconstruction depths) and 1900 non-edge, non-object pixels were selected from 213 real-world holograms captured *in situ* using the prototype holographic camera. A set of 1121 out-of-focus pixels was generated by defocusing the in-focus edges by 1.5 to 5.5 mm, randomly, in both positive and negative defocus directions. A large set of image-based features, Table 4.2, were computed for neighborhoods centered on each of the pixels, a total of 163 local features per training sample. It is worth noting that the individual focus metrics from the previous section were included as potential features.

Initial examination allowed significant pruning of the feature bank. Neighborhoods of 3×3 pixels were too small to contain enough descriptive power from a lack of statistics, while neighborhoods of 11×11 or larger also lost discriminability due to averaging; filters of 5×5 were sufficient. A number of the features reported similar conditions: for example, standard deviation of the real and imaginary components was nearly identical, and had a similar distribution as the standard deviation of the magnitude. Measures such as range and mean value could also be removed due to variations in the overall statistics for each hologram, keeping instead features which are intensity normalized.

Feature selection tools in Weka [152],[391] provide initial indicators for which of the remaining features are more likely to be discriminative for classification and detection tasks. Computing the χ^2 statistic of each individual feature for each class (in-focus, out-of-focus, and background) against the other classes suggests that the standard deviation of the gradient magnitude has surprisingly good class separability. Combinations of gradients with intensity and local standard deviation also give good χ^2 values. However, some of this sepa-

Index	Feature	Neighborhood size
1	Mean magnitude	3, 5, 7, 9, 11, 13, 15
2	Magnitude standard deviation	3, 5, 7, 9, 11, 13, 15
3	Minimum magnitude	3, 5, 7
4	Maximum magnitude	3, 5, 7
5	Magnitude range	3, 5, 7
6	Mean of real component	3, 5, 7, 9, 11, 13, 15
7	Standard dev. of real comp.	3, 5, 7, 9, 11, 13, 15
8	Mean of imaginary comp.	3, 5, 7, 9, 11, 13, 15
9	Standard dev. of imag. comp. [274]	3, 5, 7, 9, 11, 13, 15
10	Eigenvalues of Hessian matrix [153]	7
11	Eigenvalue ratio, λ_2/λ_1	7
12	Direction coher., $\frac{\lambda_1-\lambda_2}{\lambda_1+\lambda_2}$	7
13	Local histogram, seven bins	7
14	Local entropy	7
15	Hu invariants and magnitudes [164],[118]	7
16	Normalized Hu invariants	7
17	Zernike moments on magnitude [188],[63]	9
18	Zernike moments, complex field	9
19	Legendre moments [64]	7
20	Gradient magnitude, statistics	7, with $\sigma = \{1.5, 2.5\}$
21	<i>SIM</i> value	7
22	Laplacian of Gaussian	7, with $\sigma = \{1.5, 2.5\}$
23	Laplacian magnitude	7
24	Riesz transform ampl., ν, χ [352],[392]	7
25	Phase congruency [199]	7
26	Gabor filters, maximal response [198]	7

Table 4.2: Local image features computed for in-focus edge, out-of-focus edge. Features were used for automatically learning edge characteristics.

rability comes from the fact that the edge gradient scales with the hologram intensity and is affected more by overall changes in illumination than the background pixels. Using gradient features with properly normalized holograms is thus a promising concept for future work. Other features which had high χ^2 ranks included various Hu moments, Zernike moments, and Gabor filters. Ranking features based on an entropy-based information gain metric also ranked various gradient, Hu and Zernike moments as more potent choices.

The actual test for any set of features is the classification that results from their use. Initial classification tests in Weka using LogitBoost and AdaBoost models [127],[129] were computed with 35 boosting rounds, 10-fold cross validation, and decision stumps on the features. Using the entire set of image features (after removing the non-normalized statistical features, entries 1-9, 13, and 14 in Table 4.2) gave an overall recognition rate of 86.9% with 35 rounds of LogitBoosting. Using only the Zernike moments with the same algorithm gave a recognition rate of 85.1%, while the Hu moments alone had a rate of 80.5%. The individual features and select combinations from Table 4.1 had a rate of 81.7%. Results with AdaBoost were, in general, about 6-12% lower.

The Zernike moments are particularly interesting from both an image recognition and an optics standpoint. Zernike polynomials form an orthogonal set over a unit circle and are complex-valued – thus ideal for decomposing optical fields about an optical axis, as used for aberrations [36],[371]. The image moments generated from Zernike polynomials can be made shift, scale, and rotation invariant with proper scaling and re-centering of the source image [372]. The invariances and orthogonality make the Zernike moments particularly good as features for recognition (see, e.g., [188],[63],[25],[183],[372]).

Forming local features from Zernike moments was done by filtering. The Zernike polynomial, Z_{mn} , is computed over a small 9×9 pixel area and used directly as a complex-valued filter; the subscript m describes the radial order while n describes the angular order. (See Ref. [372] for the Z_{mn} equations.) Taking the absolute magnitude of the response makes the moment rotation invariant, and normalizing against the Z_{00} moment (a unit circle) removes intensity effects. Note that the usual steps for making the moments translation and scale invariant are not required; the moments are made more descriptive for edge detection by including spatial effects. A small subset of moments was found to be adequate for describ-

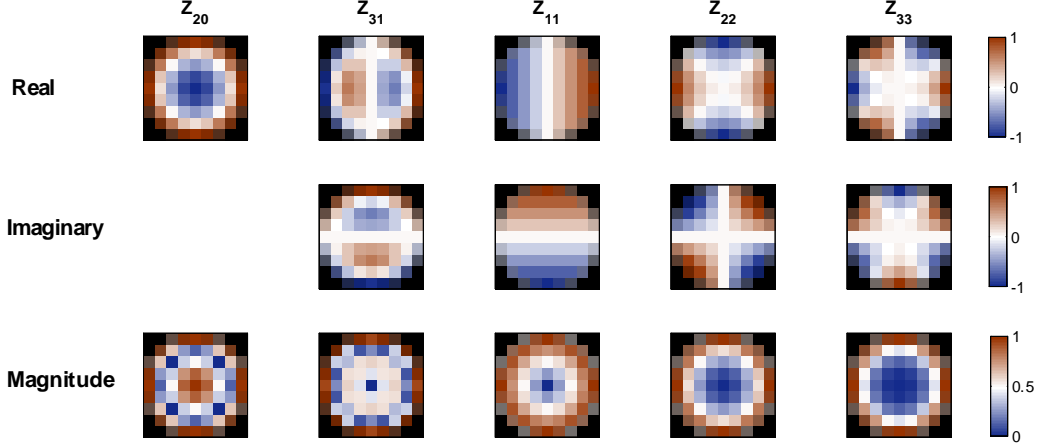


Figure 4-11: Five low-order Zernike polynomials selected as local image features. The top row shows the real part, the middle the imaginary part, and the bottom is the magnitude. (The Z_{20} filter is real-valued.) The polynomial is defined only within a unit circle; the area outside the circle is shown as black and is replaced by zeros in the image filter.

ing edges and are depicted in Figure 4-11. Interestingly, the Z_{20} moment shares similarities with a LoG filter, while the Z_{31} and Z_{11} moments are have similarities to gradient filters. The Z_{22} and Z_{33} filters both describe higher-order symmetries.

Two classifiers were trained for use as combined focus metrics and detectors. The first uses in-focus edge points as positive samples and non-object background points as negative samples, termed CZ_D here, where the D refers to object *detection*. The second classifier used in-focus edge points as positive examples and both out-of-focus edges and background samples for the negative class. This classifier is termed CZ_F here, using F to denote that it specifically detects *focused* pixels. Both classifiers make use of LogitBoosting, which uses logistic regressions on the erroneously labeled samples as the classifier is built [127],[129].

Training rates are estimated for the two classifiers by randomly removing 40% of the training data for testing and creating prototype classifiers with the remaining data. The test data is then classified and truth rates are calculated. Repeating this procedure gives statistical limits for how well the final classifier is expected to perform with error estimates. The training rates are shown in Figures 4-12 and 4-13, using 80 repetitions to estimate the errors. The training rate for CZ_D suggests that 200 rounds of boosting are sufficient, while CZ_F has possible gains beyond 300 rounds. The CZ_F has lower truth rates due to

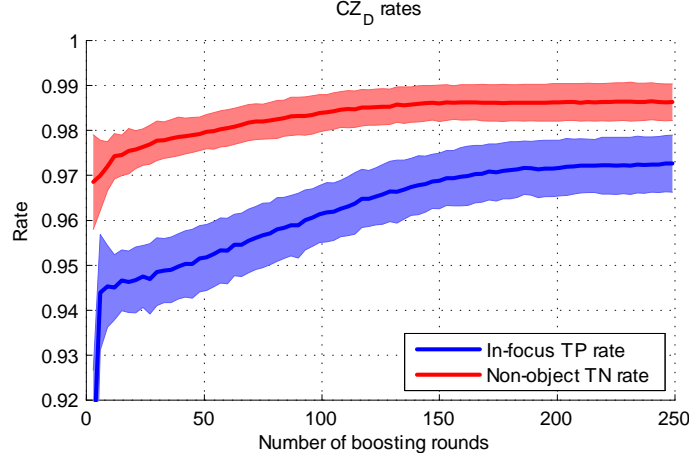


Figure 4-12: Correct classification rates for the CZ_D Zernike-based detection classifier as a function of the number of boosting rounds. Shaded regions indicate a $\pm 2\sigma$ uncertainty.

the inclusion of only slightly defocused edge samples in the negative training class, leading to greater feature overlap with the positive class and low separability.

Boosting is particularly well-suited for classification on image pixels. The decision stump model used here,

$$Fx = \sum_i \begin{cases} a_i + b_i, & f_i > \theta_i \\ b_i, & \text{else} \end{cases} \quad (4.19)$$

$$= \sum_i b_i + \sum_i a_i [f_i > \theta_i] \quad (4.20)$$

computes a classifier value Fx^3 at each pixel by adding a value $a_i + b_i$ to the vote if the feature f_i is greater than a decision threshold θ_i at boosting round i , or adding only the baseline b_i if the feature is not above the threshold. The thresholding operation makes it easy to compute Fx across entire images of features on either a CPU or a graphics card⁴. The final classification is determined by the sign of Fx : a positive value indicates support for

³Boosting was used previously in Chapter 2 for rockfish detection. The classifier there was $F(x)$, where F was the indicator function and x was a feature vector for a particular sample. The nomenclature has been changed here so that Fx represents the final output value of the classifier – and is not necessarily the function. In other words, $F(\vec{x} = \vec{f}) = Fx$.

⁴The thresholded summation of decision stumps is simple for pixel-wise GPU implementation. Alternately, a look-up-table of the sum of a_i for each feature can be created based on the thresholds, and interpolated quickly using GPU texture mapping. See Section 4.2 for discussion and details.

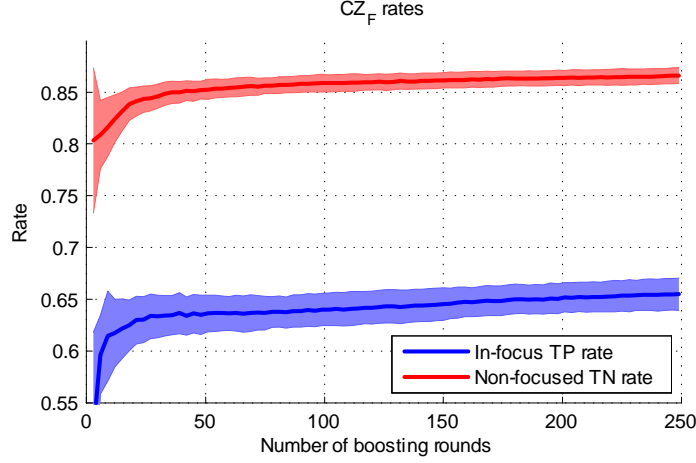


Figure 4-13: Correct classification rates for the CZ_F Zernike-based focus classifier as a function of the number of boosting rounds. Shaded regions indicate a $\pm 2\sigma$ uncertainty. The rates are significantly lower due to the inclusion of unfocused edge samples, which share similar features with edges.

the positive training class, while negative values indicate the negative class. Classification thresholds other than zero can be used to adjust for non-equal error sensitivities. (See Section 2.2.2 and Equation 2.6.) The magnitude of Fx serves as a convenient indicator for the strength of the classification and its belief level.

An example of applying the Zernike moment classifiers to a hologram is shown in Figure 4-14. The CZ_D gives strong and uniform responses around objects in this hologram, but with a wider spatial extent. The CZ_F classifier has better localization properties but with greater ambiguity around the small, point-like objects. This is a result of training the classifiers based on edges, so that particles are less likely to have strong responses.

A natural extension to the Zernike moments method for holography is to include information about the appearance of in-focus and defocused edges as those edges defocus, essentially using a set of 3D features for classification. The same five local Zernike moments as used with the CZ_D and CZ_F classifiers were computed at planes ± 2 mm and ± 1 mm from the in-focus edge and concatenated to create a 15-element feature vector describing a set of planes around the focus. This was repeated for the out-of-focus edge points and the non-object points. Initial tests with Weka showed an overall classification accuracy of 85.2% for the multi-plane Zernike moment features using 35 rounds of boosting with decision stumps

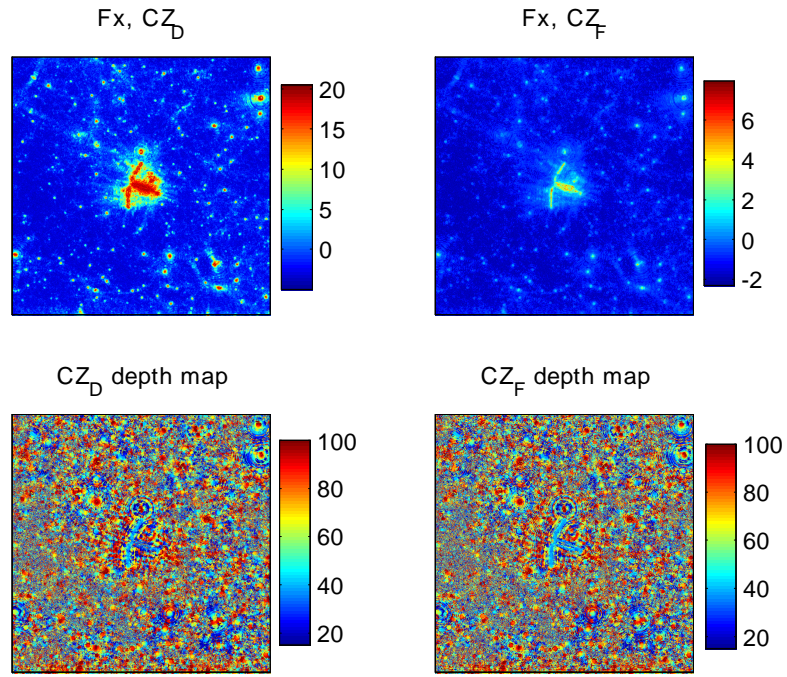


Figure 4-14: The CZ_D and CZ_F boosted classifiers applied to the sample hologram of a copepod. The maximum value of Fx at each pixel is recorded and shown in the top row. The corresponding depth maps are shown in the bottom row.

and the LogitBoost algorithm. This is marginally better than the accuracy obtained for Zernike moment features computed only at the in-focus plane – so that the additional features do not add significant discriminability for these amounts of defocus. Other classifiers (AdaBoost with 35 rounds, naïve Bayes, and SVMs) had lower performance rates, 6 to 13% worse. Given the computational burden of multi-plane features, improved 3D features are better left to future work.

Depth estimation from the Zernike moments is similar to the approach as described for the gradient-based metrics. The CZ_D result is thresholded to create a map of possible objects, and CZ_F provides a strong edge detection map. The edges corresponding to a single object are then used to estimate the depth using a weighted average where the CZ_F Fx value provides the weighting. Expected depth errors for the same synthetic hologram as Figure 4-10 were computed for the Zernike-derived depth maps. The best expected depth estimates came from using a weighted average of the pixels with the highest Fx responses, and ranged between 1.7 and 2.2 mm across SNRs between 0.1 and infinity (no added noise). This is comparable or better than the gradient-based metrics, especially considering the consistency of the result. One reason is that CZ_F includes defocused sampled points, so that it is less likely to be fooled by diffraction effects such as those of Figure 4-9. Training using a wide variety of real-world holograms with noise also helps make the Zernike method more robust. However, it should be noted that, like the gradient-based metrics, appropriate thresholds are still critical to achieving good performance – but that due to the normalization of the Zernike moments, estimating reliable and consistent thresholds may be easier in practice.

4.1.4 Object detection from raw holograms

Detecting objects in holograms by reconstructing every slice and applying a focus metric can be computationally and time intensive. Consider applying a steerable gradient filter: a minimum of five Fourier transforms (two forward, three inverse) are required to reconstruct a slice and compute the filter using frequency-domain methods. Prior knowledge about the location of interesting objects in holograms could allow for a more judicious use of computational resources and thus faster overall processing. Knowing the lateral positions of objects also helps guide depth estimation techniques. This section discusses how knowledge

about the lateral position of unknown objects can be estimated directly from the holograms without performing reconstructions.

Li and Milgram proposed a novel solution to the detection problem by reconstructing the entire volume of interest simultaneously. They noted that since a convolution is a linear operator, the sum of a number of reconstructions through a volume can be computed by convolving with the sum of the reconstruction kernels as

$$a_1 + a_2 + \dots + a_k = I * h_1 + I * h_2 + \dots + I * h_k \quad (4.21)$$

$$= I * (h_1 + h_2 + \dots + h_k), \quad (4.22)$$

where k is an index and I is the recorded hologram. Since the Fourier transform is also a linear operation, the Fourier transforms of the kernels can also be summed and used as a single kernel in the deconvolution in the same way that a single optical field is reconstructed. The sum of the kernels is a geometric series with a closed form solution [217],[254],[95], termed a “summation kernel”⁵ (SK). As more slices are added to the SK , it approaches an integral,

$$SK = \int_{z_a}^{z_b} \exp(i\pi\lambda z\rho^2) dz, \quad (4.23)$$

$$= \frac{\exp(i\pi\lambda z_b\rho^2) - \exp(i\pi\lambda z_a\rho^2)}{i\pi\lambda\rho^2}, \quad (4.24)$$

using the Fresnel approximation and setting the depth to range from z_a to z_b . The denominator specifies the main action of this filter: the ρ^{-2} dependence serves to low-pass filter the hologram. The result is that small objects with considerable high frequency content tend to disappear against the background, while large objects with low frequency content appear as large blobs.

Low-pass filtering for localization makes sense from a Wigner-domain perspective. The propagation matrix in the Wigner domain, Equation 3.59, shows that signal energy at (x, ν) is transferred to $(x', \nu') = (x + \lambda z\nu, \nu)$. Spatial frequencies with low $|\nu|$ thus remain close to x despite propagation, and selecting for these frequencies gives a good indication of the

⁵The summation kernel is also lovingly referred to as the “superkernel”.

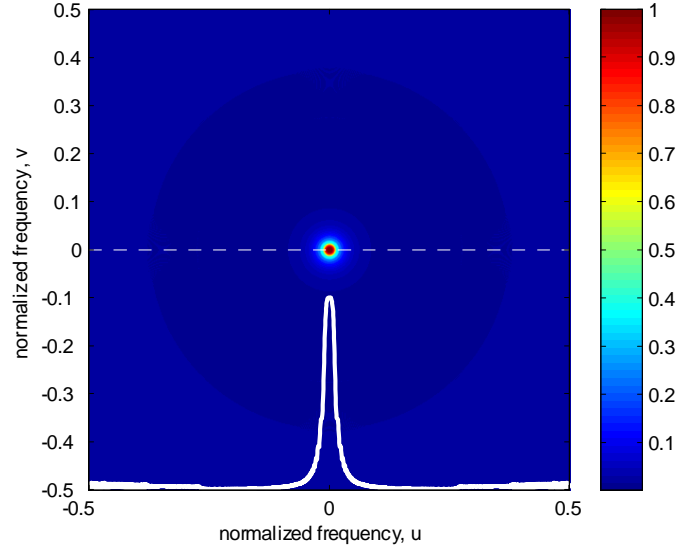


Figure 4-15: Magnitude of SK_{sph} using the depth parameters of the prototype unit. A trace of the magnitude through the center (along the dashed line) is plotted along the bottom.

reconstruction position.

The summation kernel is modified for spherical references by replacing the reconstruction depth with its planar equivalent z_{eq} (Equation 3.65),

$$SK_{sph} = \int \exp(i\pi\lambda z_{eq}(z; z_d) \rho^2) dz = \int_{z_a}^{z_b} \exp\left(i\pi\lambda\rho^2 \frac{zz_d}{z_d - z}\right) dz, \quad (4.25)$$

where $z_d = z_s + z_p$ and is the total distance from the detector to the spherical reference source point. Evaluating the integral gives

$$SK_{sph} = \left[(z_d - z) \exp\left(i\pi\lambda\rho^2 \frac{zz_d}{z_d - z}\right) - i\pi\lambda\rho^2 z_d^2 e^{-i\pi\lambda\rho^2 z_d} \Gamma\left(0, \frac{-i\pi\lambda\rho^2 z_d^2}{z_d - z}\right) \right]_{z_a}^{z_b}, \quad (4.26)$$

where Γ is the incomplete Gamma function and stems from the exponential integral [12],[1]. Computing the integral numerically, plotted in Figure 4-15, demonstrates that SK_{sph} has a similar low-pass filtering effect as SK . The differences are that the peak is slightly narrower in SK_{sph} than SK due to the larger equivalent reconstruction depths, and there is less ringing due to the non-linear increase of z_{eq} with z so that the phase variations average out.

Detecting objects from the SK filter is challenging due to the variability in the response strength and size. Extremely low-frequency fluctuations in background intensity can be larger than peaks from small objects so that direct thresholding removes all but the strong peaks in the high intensity regions. Using local statistics to calculate a z -score for peaks works well when the size of the local neighborhood is larger than the object of interest but smaller than the mean separation, so that the objects do not dominate the statistics. Selecting an *a priori* local neighborhood size for point-like particles is possible, but is in general not an option for objects of unknown size.

A multi-scale approach is preferred instead. Lindeberg notes that the difference of local statistics at different scales is an approximation for the derivative of a scale-space pyramid in the scale direction (and an approximation to the diffusion of the original image over different scales). That observation leads directly to the use of a scale-normalized Laplacian, a derivative operator, for higher-accuracy blob detection [223]. The image of interest, in this case the SK response, is smoothed at multiple scales using a Gaussian, then filtered with a Laplacian and the response normalized to the scale. Extrema in the scale-space give the peaks⁶ at appropriate scales [223],[224].

It is interesting to note that the Laplacian,

$$B = (\partial_{xx} + \partial_{yy})(sk * G(\sigma_s)), \quad (4.27)$$

where sk is the hologram response to the SK filter and $G(\sigma_s)$ is a Gaussian with scale σ_s , is the trace of the Hessian of the Gaussian-smoothed SK response. The Hessian is positive definite for minima and negative definite for maxima, so that its eigenvalues have the same sign. The trace is then expected to have a larger magnitude (when the signs of both eigenvalues agree) near the extrema. Thus, the Laplacian blob detector seeks scale-space peaks of the smoothed SK response. It is important to recognize that the trace has a larger magnitude for isolated, round peaks, where both Hessian eigenvalues have large magnitudes, than for peaks with longer extent in one direction, where the Hessian eigenvalue in the direction of the extension is small. Thus, the Laplacian can be expected

⁶“Peaks” refers here to both positive and negative isolated responses. A raw hologram results in dark, negative blobs, while a background-subtracted hologram gives bright blobs in the magnitude image.

to detect long blobs, such as from rod diatoms and filaments, with a lower response than for round objects.

A second approach to object detection stems from examining the frequency landscape around an object. The summation kernel specifically seeks areas with low-frequency content. In addition to the low frequencies near the object, there should be higher frequencies, ν , at distances $z\nu$ from the scattering point. The local frequency landscape should then dip around an isolated object.

Computing the 2D local frequency to search for dips is surprisingly difficult. The Wigner transform is one obvious choice, with the dominant local frequency selected by taking the maximum along the ν direction; finding the transform for a large hologram, however, is too computationally expensive to be feasible, $O(N^4)$, where N is the pixel count of one side. Similar methods, such as peak picking of an S -transform [359],[360],[282] or of a local Fourier transform (analogous to a spectrogram with a sliding window) are considerably slow to calculate, extremely limited in frequency resolution, or have poor results; a local Fourier transform is itself $O(N^4)$, though the reduced size of the window makes it significantly faster than a full Wigner transform. Using a series of bandpass filters and retaining the frequency with the maximum response is $O(N^2)$ using Fourier transforms – but also gives poor results, often dominated by low spatial frequencies. Wavelets are worth mentioning as methods for local frequency estimation, also $O(N^2)$, but the appropriate choice of wavelet and interpolation of a dominant frequency from amongst multiple scales is challenging.

One recent alternative comes from extending the Hilbert transform to higher dimensions. The Hilbert transform is used in 1D to find the analytic signal, $f_a(x) = A(x) \exp(i\phi(x))$, which generates the measured signal, $f(x)$, by taking the real component of f_a . The resulting analytic signal gives both an envelope amplitude, $A(x)$, and an instantaneous phase, $\phi(x)$, the latter of which can be differentiated to find the local frequency [114]. Defining the Hilbert transform for higher dimensions is not obvious, namely because the Hilbert transform divides the Fourier space into positive and negative frequencies – and the sign of the frequency is ambiguous for two or greater signal dimensions [174]. Larkin et al. and

Unser et al. propose using a frequency domain filter,

$$R(u, v) = \frac{u + iv}{|u^2 + v^2|^{1/2}} = e^{i\theta(u,v)}, \quad (4.28)$$

where u and v are the spatial-frequency domain variables, or θ is the polar angle. This has the unique property that a slice through the center of the kernel at any angle is anti-symmetric, just as the Hilbert transform is anti-symmetric about the frequency origin [211],[210],[392]. (It is also interesting to note that this kernel is equivalent to a spiral phase plate with a single twist, such as used for generating optical vortices; see, e.g. [196].) The filter that Equation 4.28 implements is known as the Riesz transform, a vector-valued Hilbert transform [352]. The signal generated from R operating on f is \hat{f} , and is the analogue of the imaginary component of the analytic signal – though in this case, \hat{f} contains both real and imaginary components itself. The amplitude envelope is then

$$A(x, y) = \sqrt{f^2(x, y) + |\hat{f}|^2(x, y)}, \quad (4.29)$$

and the instantaneous phase is given by

$$\phi(x, y) = \arctan\left(|\hat{f}|/f\right). \quad (4.30)$$

The local frequency is the directional derivative of $\phi(x, y)$ in the direction perpendicular to the fringe and can be found from f and \hat{f} ; the details of this calculation are discussed with more detail in [392] for the interested reader (see also their Appendix I). Unser et al. also provide a directional coherency measure,

$$\chi = \frac{\lambda_{\max} - \lambda_{\min}}{\lambda_{\max} + \lambda_{\min}},$$

where the λ are the eigenvalues of the local weighted structure matrix, itself a local version of a Hessian. Thus, areas with strong fringes and good orientation estimation have $\lambda_{\max} \gg \lambda_{\min}$ and $\chi \simeq 1$; similarly, areas with weak fringes have $\lambda_{\max} \simeq \lambda_{\min}$ and $\chi \simeq 0$. The χ metric provides a rough measure of the reliability of the local frequency estimate [392].

A wide array of techniques beyond the Hilbert transform have been developed for fre-

quency estimation of 1D signals (see, e.g., [148],[294],[293],[213],[341],[342],[35]). These can be utilized by interpolating a 1D signal perpendicular to the fringe direction in areas of high directional coherency. Fringe orientation and χ are available from the Riesz method; alternate orientation estimators are discussed in [212] and [126]. The Quinn-Fernandes (QF) iterative method is used here for its speed, excellent frequency resolution, and accuracy [294],[293] with only limited data points (and thus a highly localized area). A 2D frequency estimate is calculated from the 1D QF estimate by taking an average of the QF frequencies weighted by χ and by a Gaussian distance.

Comparisons of four local frequency estimation methods are shown in Figure 4-16 for sample holograms. The first example, a synthetic chirp signal, has excellent frequency estimation from bandpass filters, the maxima of 2D spectrograms, and the QF interpolation. The local frequency derived from the Riesz transform has low directional coherence in the center of the pattern, then begins ramping up linearly away from the center. As the chirp reaches high frequencies, the Riesz-estimated frequency actually decreases. This is due partly to implementation: derivatives are taken using steerable gradients over a local neighborhood, so that there is low-pass filtering embedded into the processing chain. Natural objects have lower amounts of high-frequency signal, so this is less of an issue for holograms C and D. There the Riesz actually provides a more unique signal for detection than the spectrogram peaks or QF methods and is computed significantly faster.

Sample detections for *in situ* holograms captured using the prototype system are shown in Figure 4-17. A multiscale detector was implemented and used with both the *SK* response and the Riesz local frequency. Holograms A and B have small objects distributed through the volume, many with nearly symmetric diffraction rings. The filament to the right side of hologram A is detected by only one strong blob (#7), as expected from its smaller Hessian trace. Similarly, the filaments in hologram E are each detected by multiple blobs along their length. Detections using the Riesz local frequency landscape are similar to *SK*. One notable exception is the filament in the bottom right of hologram E, where *SK* finds an object within the bent corner (a coherent diffraction effect), towards which the Riesz frequency shows no positive response. Hologram F does not contain any objects of interest, and only noise and unidentifiably small particles are detected.

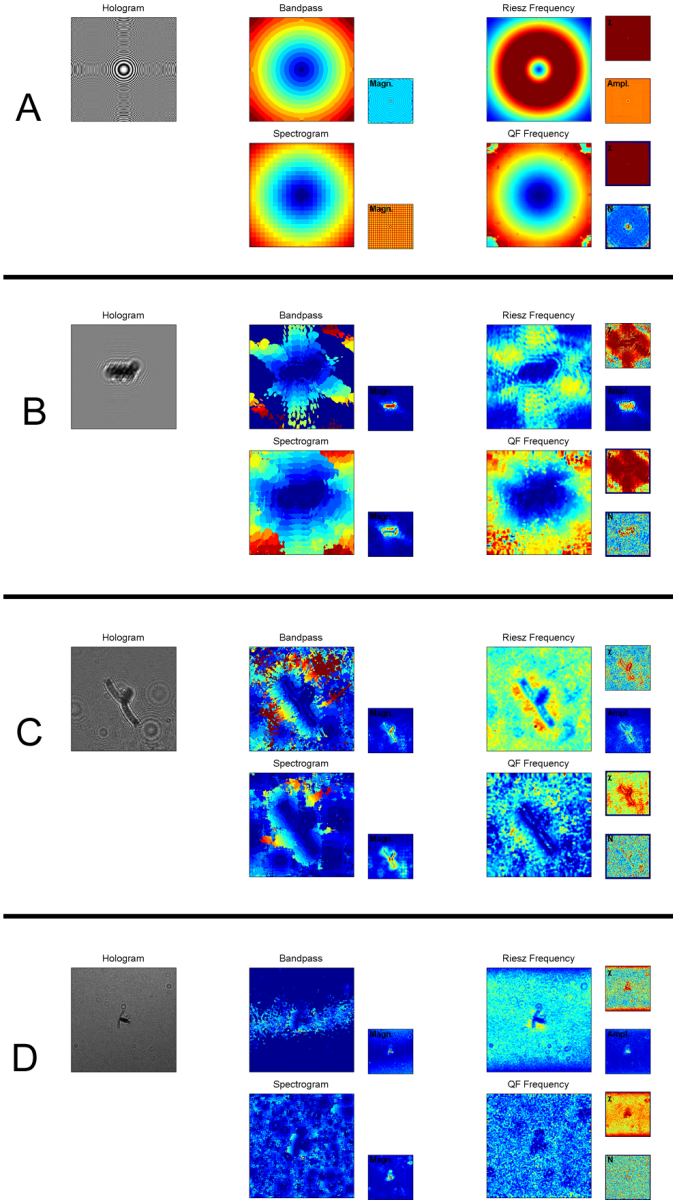


Figure 4-16: Application of local frequency estimation to sample holograms. Hologram A is a chirp signal with linearly increasing frequency, B is a synthetic hologram, and C and D are holograms of two copepods. The frequency estimated using bandpass filters and the magnitude are shown in the center top panels. A sliding window and its peak magnitude method is shown in the center bottom panels. The local frequency from a Riesz transform, its directional coherency χ , and the signal amplitude are in the top right panels. The frequency using a Quinn-Fernandes iterative method, along with the χ values at the sample points and the number of QF iterations, N , are in the bottom left panels. The samples all use the Matlab *jet* colorscheme, with blue as the minimum and red as the maximum. Plots are scaled between the minimum and maximum values, except for χ , which ranges from 0 to 1.

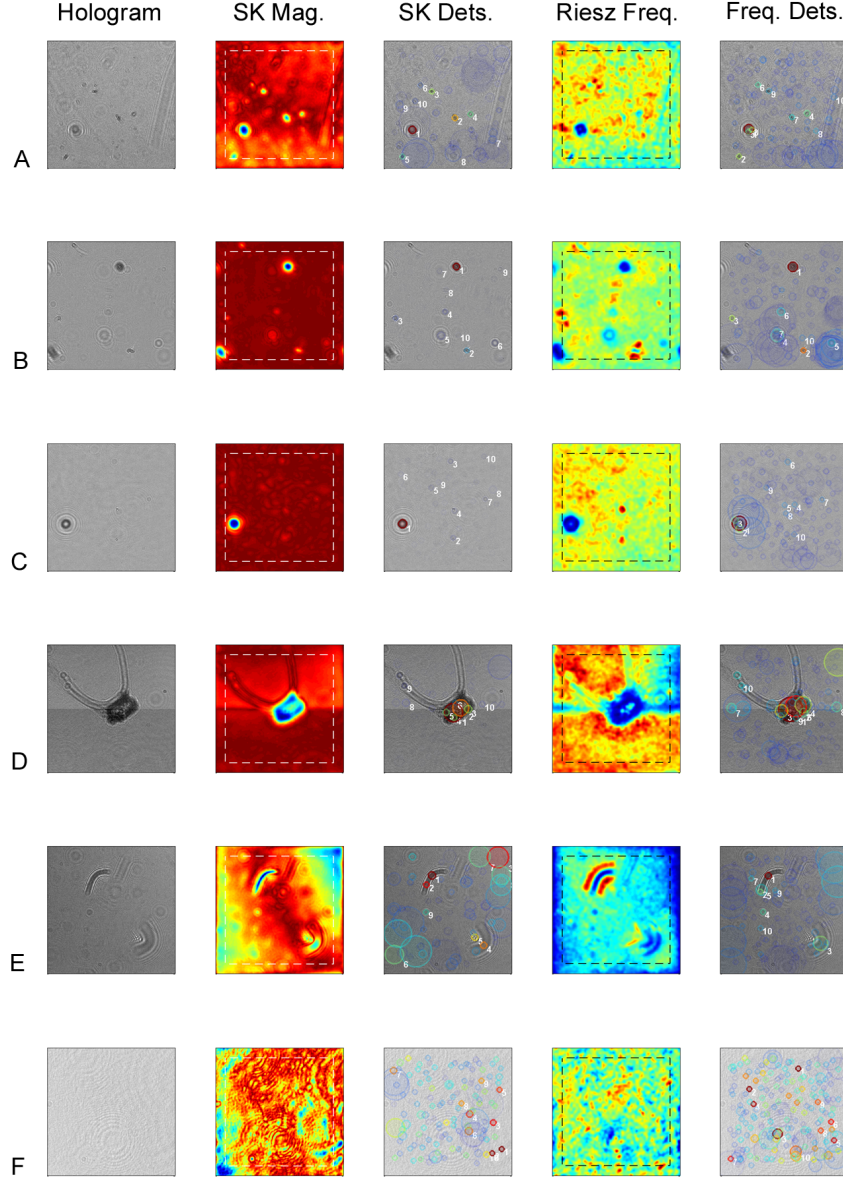


Figure 4-17: Object detections from summation kernel and Riesz local frequency landscapes. The original hologram is shown in the left-most column. The *SK* response is the second column, with strong (but negative-valued) detections in blue. Detections from the *SK* response are in the center column, labeled in increasing order with decreasing detection strength. The size of the circle indicates the scale. Color and transparency are also used as hints to the strength of the detection. Only those detections which fall inside the dotted line in the *SK* image are shown. The Riesz frequency landscape is the second column from the right, with blue denoting low frequency and red as higher frequencies. Detections for the Riesz frequency are plotted in the right-most column. Hologram C is an oil droplet. Hologram D contains a large object, approximately the size of the largest detection scale. Holograms A, D, and E contain thin, long objects. Hologram F contains no strong or interesting objects, with detections based on noise.

Detections from multiple scales are combined into a single detection here using a local extrema scheme. Clusters are formed by comparing detections from neighboring scales that fall within each others' scale size, with the smaller response joining the cluster of the stronger response. Detections which overlap multiple neighbors are grouped with the neighbor nearest, laterally. An improvement for future work would be to use a mean-shift type of scheme to combine each detection [72],[61] (with modifications for varied scale and dimensionality [71],[139]), especially for filaments and large objects such as holograms D and E. Other methods for grouping regions such as clustering and mode-seeking [106],[400],[278] or maximally stable extrema regions [243],[101],[122] are also recommended for detecting objects in the SK or Riesz landscapes, with the caution that these methods can be sensitive to selecting appropriate settings.

The SK response and the Riesz frequency give complementary information. The SK responds to local energy (and is thus related to minimum intensity type metrics), so that objects with strong absorption and low frequency content have high contrast. SK is expected to have a better performance on holograms with high noise, such as from particles or camera-readout, thanks to its low-pass filtering effect. The Riesz frequency responds instead to the dominant frequency, so that small objects with strong scattering into higher spatial frequencies should be better detected. The Riesz method is expected to perform better for holograms with low noise. Detections that appear in both responses are likely true objects and not false positives, especially in the case of low noise.

Deciding whether a candidate blob detection should be processed further depends on a number of factors. An obvious criterion is that the blob should have a “significant enough” response – which could mean that its raw response strength is high, or that its relative response is higher compared to all the detections in a hologram.

The performance of the SK and Riesz methods was measured using a selection of 236 objects from 66 *in situ* holograms marked by hand. The holograms were processed by the summation kernel and Riesz methods, and the detections compared against the known set of objects. The summation kernel had fewer total detections for the given settings, an average of 166 per hologram compared to 226 per hologram for the Riesz method. That made the SK method slightly more efficient, as fewer detections needed to be reconstructed

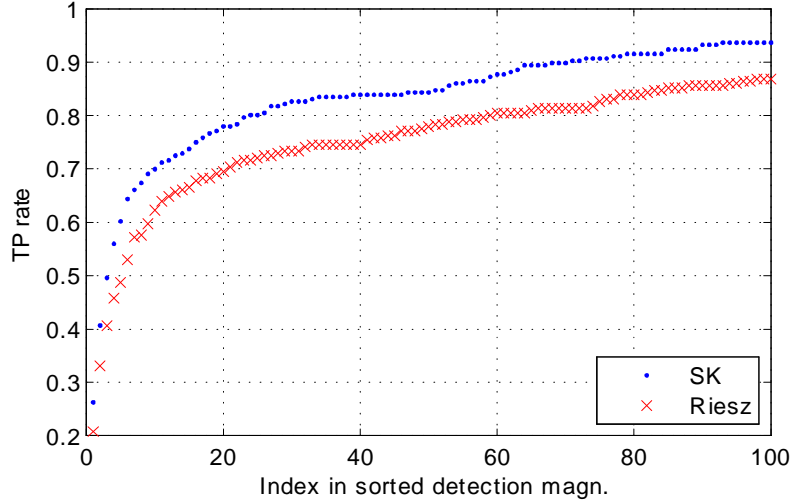


Figure 4-18: Percent of known objects found within the n strongest detections, where n is the index in the sorted detection magnitudes. The plots are the cumulative distributions for detecting objects.

to find the specific objects: 70% of the hand-marked objects were found within the first ten detections (those with the strongest relative blob responses), while the Riesz method had only 62% within the first ten (Figure 4-18). Detections can also be selected based on the raw blob response strength by setting a decision threshold. The effect of scanning the threshold is plotted in the precision-recall curves of Figure 4-19. The Riesz method has better performance for low recall, then approximately matches the *SK* method for higher recall rates. One caveat for these results is that all objects in the holograms larger than about $50\ \mu\text{m}$ were marked to measure the *SK* and Riesz methods purely for their ability to detect objects. Most plankton is larger than $50\ \mu\text{m}$, leading to larger and stronger responses for both detection methods, so that rates for biologically interesting planktonic objects is expected to be much higher.

The two detectors have differing performance based on which decision criteria is used: whether the top-ranked relative detections should be reconstructed, or if detections above a fixed absolute threshold should be used instead. This hints that a better decision for which detections to process could be determined by a combination of factors. Future work in improving the detectors and the processing decision, especially using machine learning and additional features such as the coherency or spectral energy, is suggested, especially

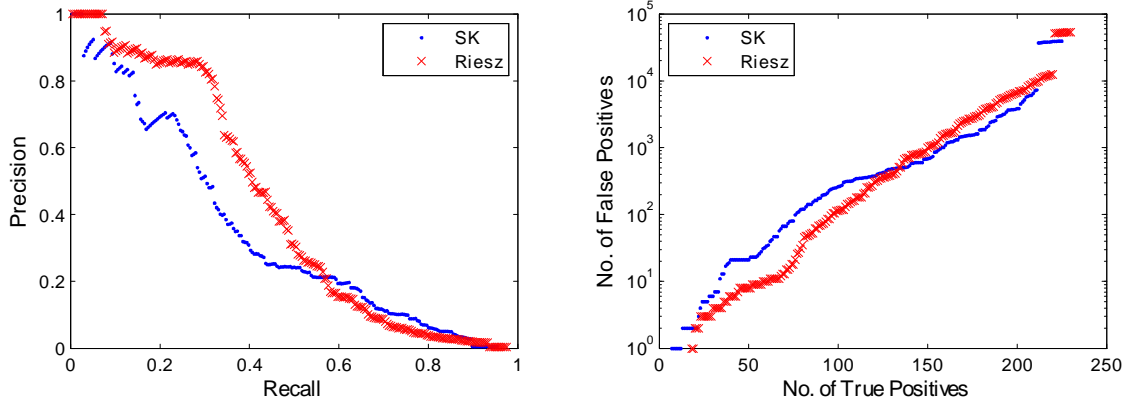


Figure 4-19: Left: Precision and recall rate as the blob strength decision threshold is scanned. The Riesz method has better precision, returning fewer false positives per true positive until the recall rate reaches about 0.55; at that point, the *SK* method is slightly better. Right: the number of true positives and false positives as the decision threshold is scanned. The total number of true positives is 236.

for noisy real-world holograms. Laboratory holograms with controlled exposures and noise conditions are expected to give much cleaner detection results.

4.1.5 Fast depth estimation

The focus measures of Sections 4.1.1–4.1.3 are applied to a set of reconstructions to decide the plane of best focus for some object. In the absence of any other knowledge, the depth of an object can be determined by reconstructing every plane in the volume. This can be computationally slow and taxing even with fast methods and processing (see Section 4.2). Interesting objects in the holograms from the prototype are rare. For example, the previous detection section (4.1.4) had an average of 3.6 objects per hologram in the test set – with the strong caveat that the holograms selected for hand-labeling had crossed an initial threshold for variability and were more likely to contain objects of interest in the first place. An *a priori* estimate of the depth of the objects could substantially reduce the number of reconstructions if only a few planes needed to be searched for the in-focus object. A limited number of depth estimators have been proposed based on the shearing properties of the Wigner transform [268],[190], but the well-known cross-talk in Wigner transforms with multiple objects and slow computation makes these methods inapplicable for all but the simplest synthetic test cases.

This section reports on recent work done by Li, Loomis, Hu, and Davis to develop a fast depth estimator [220],[218],[219], then extends the analysis of the estimator to better understand its limits. The original estimator has excellent performance for laboratory test cases but suffers for *in situ* holograms reconstructed over extended depth ranges. A new estimation metric is proposed based on insights gained from the original and tested on real holograms captured by the prototype unit to measure its effectiveness.

The information in a hologram is given by Equation 3.13 (or by Equation 3.86), so that the Fourier transform is

$$\hat{I} = A(u, v)H(u, v; z_p) + A^*(-u, -v)H^*(-u, -v; z_p), \quad (4.31)$$

where A is the spectrum of the illuminated object (or of $t(x, y)$, without loss of generality, as discussed in Section 3.2.5; $A(u, v)$ and $a(x, y)$ are used here for consistency with [220]). If the object can be approximated as real valued, the Fourier transform of A is conjugate-symmetric and

$$\hat{I} = A(u, v)[H(u, v; z_p) + H(u, v; -z_p)] = 2A(u, v)\cos\left(\frac{2\pi}{\lambda}z_pf_r\right), \quad (4.32)$$

using the properties of H and substituting $f_r = [1 - (\lambda u)^2 - (\lambda v)^2]^{1/2}$ as the radial spatial frequency. This equation shows that the spectrum is modulated by a chirped cosine based on the depth of the object, z_p . Determining the depth is a matter of estimating this chirp rate.

The approach used by Li et al. is to apply a matched filter to the \hat{I} spectrum. Specifically, since A is complex-valued, an ℓ_1 norm⁷, the sum of absolute values, is used on \hat{I} to limit the expectation to non-zero values and provide computational tractability. The

⁷The ℓ_p norms (or ℓ^p in some texts) are technically vector norms [387], while the L^p norms are reserved for functions. The derivation is illustrated on L norms for clarity, while the actual computations are performed on discrete elements with an ℓ norm.

matched filter is then

$$F_c = \iint \left| \hat{I}(u, v) \right| \left| \cos \left(\frac{2\pi}{\lambda} z_t f_r \right) \right| dudv, \quad (4.33)$$

$$= \iint \left| 2A(u, v) \cos \left(\frac{2\pi}{\lambda} z_p f_r \right) \cos \left(\frac{2\pi}{\lambda} z_t f_r \right) \right| dudv, \quad (4.34)$$

where z_t is the test distance. An associated metric using a sinusoid,

$$F_s = - \iint \left| 2A(u, v) \cos \left(\frac{2\pi}{\lambda} z_p f_r \right) \sin \left(\frac{2\pi}{\lambda} z_t f_r \right) \right| dudv, \quad (4.35)$$

will also be useful. The cosine terms of Equation 4.33 can be rewritten as

$$\cos \left(\frac{2\pi}{\lambda} z_p f_r \right) \cos \left(\frac{2\pi}{\lambda} z_t f_r \right) = \frac{1}{2} \left[\cos \left(\frac{2\pi}{\lambda} f_r [z_p + z_t] \right) + \cos \left(\frac{2\pi}{\lambda} f_r [z_p - z_t] \right) \right], \quad (4.36)$$

so that F_c becomes

$$F_c = \iint \left| A(u, v) \left[\cos \left(\frac{2\pi}{\lambda} f_r [z_p + z_t] \right) + \cos \left(\frac{2\pi}{\lambda} f_r [z_p - z_t] \right) \right] \right| dudv. \quad (4.37)$$

The first cosine term is related to the twin image while the second is related to the object coming into focus and goes to unity at $z_p = z_t$. The effect of each term on A can better be seen by examining the equivalent convolutions in the spatial domain. Noting that each of the cosines can be written as the sum of two propagation kernels, the direct inverses are

$$\mathcal{F}^{-1} \left\{ \cos \left(\frac{2\pi}{\lambda} f_r [z_p \pm z_t] \right) \right\} = \frac{1}{2} h(x, y; z_p \pm z_t) + \frac{1}{2} h(x, y; -(z_p \pm z_t)), \quad (4.38)$$

ignoring the arbitrary phase offset term. Substituting the Fresnel approximation, Equation 3.16, for h ,

$$\frac{1}{2} h(x, y; z_p \pm z_t) + \frac{1}{2} h(x, y; -(z_p \pm z_t)) \simeq \frac{\cos \left(\frac{ik}{2(z_p \pm z_t)} [x^2 + y^2] \right)}{i2\lambda(z_p \pm z_t)} \equiv c(z_p \pm z_t). \quad (4.39)$$

The result is that $c(z_p - z_t)$ dominates over $c(z_p + z_t)$ due to the divisor in the kernel when

z_t is near z_p , and

$$a(x, y) ** (c(z_p + z_t) + c(z_p - z_t)) \simeq a(x, y) * c(z_p - z_t). \quad (4.40)$$

Thus, near the focus,

$$F_c \simeq \iint \left| A(u, v) \cos \left(\frac{2\pi}{\lambda} f_r [z_p - z_t] \right) \right| dudv, \quad (4.41)$$

which takes a maximum at $z_p = z_t$. Similar arguments for F_s show that

$$F_s \simeq - \iint \left| A(u, v) \sin \left(\frac{2\pi}{\lambda} f_r [z_p - z_t] \right) \right| dudv. \quad (4.42)$$

Computational experiments strongly suggest that the twin image component is not significant over a large range of z_t [220]. Both F_c and F_s should maximize at the object distance, so that $F_c F_s$ provides a simple form of voting between the two estimates and a cleaner response. Taking the product also eliminates some issues with harmonics (F_c and F_s have opposite signs at $z_t = z_p/2$, for example).

The F_c and F_s measures are integrated over the frequency domain for each test depth. The computation can be reduced significantly by noting that f_r in the cosine and sine terms is radially symmetric. Rewriting Equation 4.33 in a polar coordinate system gives

$$F_c = \int_0^{\varrho} \int_0^{2\pi} \left| \hat{I}(f_r, f_\theta) \right| \left| \cos \left(\frac{2\pi}{\lambda} z_t f_r \right) \right| f_r df_\theta df_r, \quad (4.43)$$

$$= \int_0^{\varrho} \left[\int_0^{2\pi} \left| \hat{I}(f_r, f_\theta) \right| df_\theta \right] \left| \cos \left(\frac{2\pi}{\lambda} z_t f_r \right) \right| f_r df_r, \quad (4.44)$$

where $f_\theta = \arctan(u, v)$ is the angle in the spatial-frequency domain and ϱ is the maximum radial spatial frequency. The inner integration needs to be computed once per hologram. The matched filter response can then be computed by integrating over f_r for each z_t , reducing the focus measure to a single integral. This approach of using a polar decomposition to quickly compute the responses is named the *PFF* method (“polar frequency focus”).

One of the major caveats of the PFF is that it has cross-talk for multiple objects. Up to

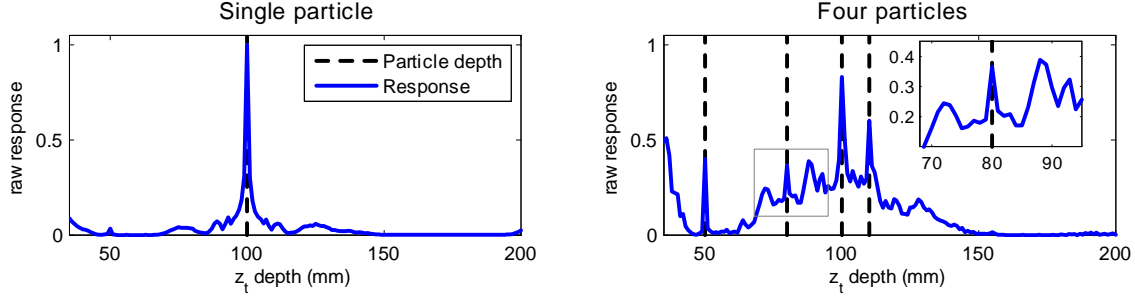


Figure 4-20: Cross-talk between multiple objects. Left: a single particle with a sharp peak the $F_c F_s$ response. Right: four objects with decreased peak contrast and spurious smaller peaks due to cross-talk. The particle peak at $z_p = 80$ is actually smaller in magnitude than the cross talk at $z_t \simeq 88$ mm (inset).

this point, the discussion and derivation has centered on single objects with the assumption that the objects will be spatially rarefied enough so that the portion of the hologram used to compute the spectrum is dominated by the spatial frequencies from only the object of interest. The derivations become less clear as more objects are added. For a small number of additional objects (so that the $\sum_{j,k} o_j o_k^*$ cross-talk terms are negligible), the F_c estimate becomes

$$F_c \simeq \iint \left| \sum_k A_k(u, v) \cos \left(\frac{2\pi}{\lambda} f_r [z_{p,k} - z_t] \right) \right| dudv, \quad (4.45)$$

where A_k and $z_{p,k}$ correspond to the k^{th} object. If the k objects have well-separated spectra, the sum is nearly independent and F_c behaves like a sum of the two responses. However, when the A_k (and their associated depth modulations) are not independent, the absolute value of the sum is harder to predict. Figure 4-20 illustrates the cross-talk issue between small simulated circular particles. Each particle has a similar size and depth, so that the spectra are not well separated. Note that the four particle case includes a spurious cross-talk peak in the $F_c F_s$ product that is actually larger than the response peak from a nearby particle.

An ℓ_1 norm was initially chosen for its analytical tractability and fast computation. The ℓ_1 norm, as Li notes, is also a good approximation for an ℓ_0 sparsity norm. Other norms

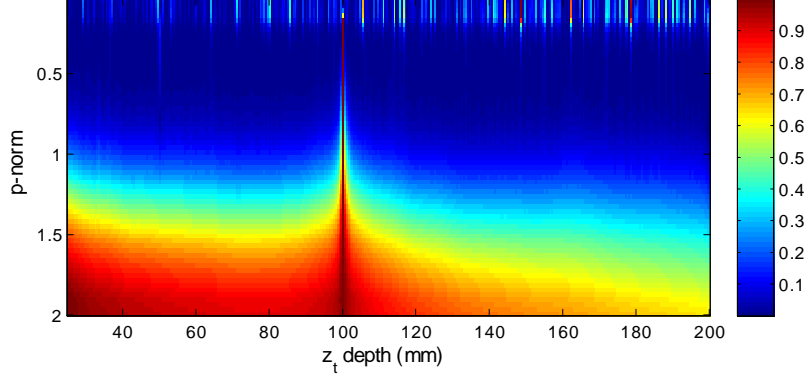


Figure 4-21: Computed responses for various p -norms for a synthetic copepod (the fifth object of Figure 4-22) located at $z_p = 100$ mm. The response is normalized to a maximum of 1.0 for each p -norm. Higher ℓ_p have less detection contrast, while too low of an ℓ_p leads to high numerical noise; a “sweet spot” for this object appears to exist around $p = 0.45$.

are of course possible. For an arbitrary L^p norm,

$$F_{c,p} = \left[\iiint \left| \hat{I}(u, v) \right|^p \left| \cos \left(\frac{2\pi}{\lambda} z_t f_r \right) \right|^p dudv \right]^{1/p}, \quad (4.46)$$

with the discrete summation performed in ℓ_p , and similarly for $F_{s,p}$. The analysis for $p = 2$ is particularly elegant and leads to convenient relationships via the Cauchy-Schwarz inequality. (Derivations for specific norms are academic in nature and follow similar approaches as for the ℓ_1 . In practice, the performance is not tied to an exact norm, varies from the analysis, and depends upon the implementation chosen, e.g., removal of background or normalization. Thus, the analysis for other norms is omitted here for brevity.) Figure 4-21 shows the $F_{c,p}F_{s,p}$ response for different p -norms and a synthetic copepod hologram. The contrast decreases for higher p , while low p has larger amounts of numerical noise.

The response of the PFF is difficult to predict in closed form. Instead, a number of simulations were computed with synthetic holograms to verify its performance under various conditions (see also [220] and results presented therein). The PFF tends to give the sharpest response peaks for small objects and those whose spectra are not radially symmetric. Gaussian noise in the hologram adds an offset to the spectra but does not significantly affect peak detection until the SNR drops below about 5. The response is consistent with depth, expanding slightly like $z_p^{1/2}$ for a point source and providing extremely good depth

estimates (depths within 0.5 mm for a point source located between 50 and 200 mm with a 1024×1024 point spectrum). These results would indicate that for sparse object distributions where only a few objects are present in the hologram, the PFF provides a fast and accurate depth estimator.

The actual performance of the PFF for real-world *in situ* holograms was tested using 174 hand-marked objects. Each of the holograms was captured with the prototype imager. A 1024×1024 pixel window was extracted around the known objects, the power filter applied (Section 3.3.4), and the PFF computed. The ensemble mean at each z_t was calculated and smoothed to estimate the bias due to particle noise, then used to normalize the individual responses to the same apparent background level. Peak detection is done by smoothing the responses with a series of Gaussians with increasing width, marking peaks which are consistent across scale (similar to the multi-scale object detection in Section 4.1.4). The positions of the peaks and the response strength at the optimal scale were extracted, and the candidate peaks ranked by the magnitude of the response.

The statistical data for real-world holograms is, unfortunately, far from satisfying. Higher numbers of competing objects, low spectral energy of target objects, particles, noise, and a large depth range all make the object peaks small compared to the background. The majority of the hand-marked objects did not correspond with the strongest peaks; on average, the objects corresponded with peaks 42% down the ranked peak list. (The best was seen at $p = 1.5$, an average of 37% down the ranked list. Norms from $p = 0.25$ to $p = 4$ were tested.) The expected error between the known positions and the nearest peaks was consistent with the error if peaks were randomly distributed over the z_t depth range. The conclusion is again that the PFF is suitable for extremely limited numbers of objects within the holograms, decidedly not the case with marine holograms captured with the prototype device.

The PFF peaks when the chirp terms in the object spectrum and the test filter match. One of the problems is that this matching is weighted by the spectrum, A . A few strong peaks in the spectra can thus affect the overall depth matching. Similarly, since the response strength is proportional to $\iint |A|^p du dv$, an object with weak spectra content can easily disappear into the background noise. Given that the purpose of the PFF is to estimate

depth by comparing the chirp terms, taking steps to reduce the effects of the object spectra should improve the results. The approach taken here is to compute a related focus metric on a normalized spectrum,

$$M_c(z_t; p) = \int_0^{\varrho} \frac{1}{\widehat{I}_\theta(f_r; p)} \left[\left| \cos\left(\frac{2\pi}{\lambda} z_t f_r\right) \right|^p - \mu_p \right] \int_0^{2\pi} |I(f_r, f_\theta)|^p df_\theta f_r df_r, \quad (4.47)$$

$$= \int_0^{\varrho} \left[\left| \cos\left(\frac{2\pi}{\lambda} z_t f_r\right) \right|^p - \mu_p \right] \left| \cos\left(\frac{2\pi}{\lambda} z_p f_r\right) \right|^p \quad (4.48)$$

$$\times \frac{1}{\widehat{I}_\theta(f_r; p)} \int_0^{2\pi} |A(f_r, f_\theta)|^p df_\theta f_r df_r \quad (4.49)$$

where μ_p is the mean value of $|\cos(2\pi z_t f_r / \lambda)|^p$ and serves to reduce background bias (making it closer to a true matched filter), and $\widehat{I}_\theta(f_r; p)$ is an estimate of the spectral weighting and is generated from the f_θ integral. (Even in cases where the ensemble spectra can be estimated *a priori*, the variations between specific samples is enough to warrant a per-hologram estimation.) The $\int |A|^p df_\theta$ term is assumed to vary slowly with f_r and provides a limiting envelope, while the cosine term oscillates quickly for middle- to high-frequency f_r and z_p in the range of tens to hundreds of millimeters. The magnitude of the Hilbert transform naturally provides an estimate of the $|A|^p$ envelope. However, the Hilbert transform gives poor estimates when the frequency of the chirp is low. Numerical experiments show that estimating $|A|^p$ using a large moving average filter provides a better \widehat{I}_θ normalization function. A paired focus metric, M_s , replaces the z_t cosine with a sine. Like F_c and F_s , the two M_c and M_s give similar and complementary results. However, since M_c and M_s can both be zero or negative, the sum of the two is the more appropriate combination. This second approach of a normalized spectrum and matched filtering on the chirp will be called the MPFF⁸.

The MPFF, like the PFF, has a behavior which intuitively makes sense but is difficult to predict in a closed form. A number of numerical experiments are presented to provide an idea of what to expect. Example responses of $M_c + M_s$ are shown for synthetic amplitude

⁸The approach is also termed the “1-4-5” depth estimator due to a particular parameter set.

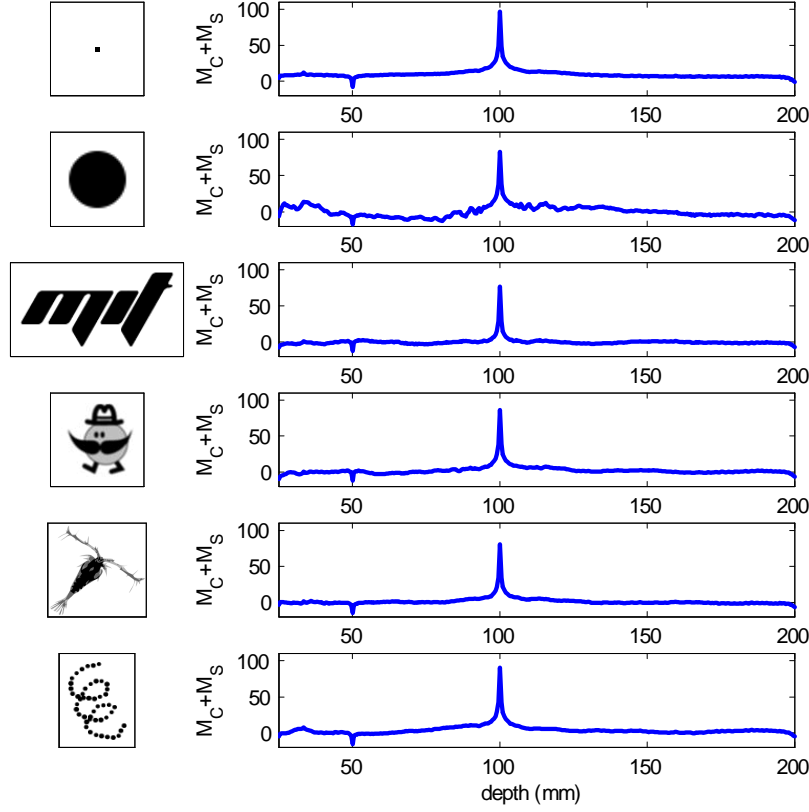


Figure 4-22: Focus responses for sample objects. The object is shown to the left and the sum $M_c + M_s$ is plotted to the right. Each object was embedded in a 1024×1024 pixel array. The correct depth is $z_p = 100$ mm in each case.

objects in Figure 4-22. Holograms at $z_p = 100$ mm were created for each object and the focus metric computed. In each case, the response has a clean peak at $z = 100$ mm and a low background level.

The $M_c + M_s$ responses for a simulated point source are shown in Figure 4-23. The point source was scanned over a range of 50 to 225 mm and the MPFF used to estimate the depth by searching for the maximum response. The width of the response grows with z_p , but the peak occurs within 0.05 mm of the actual depth. The hologram here is 1024×1024 pixels and each response vector has been normalized to unit length for comparison of the energy spread at higher z_p .

The MPFF has excellent noise characteristics. For example, consider a noise spectrum,

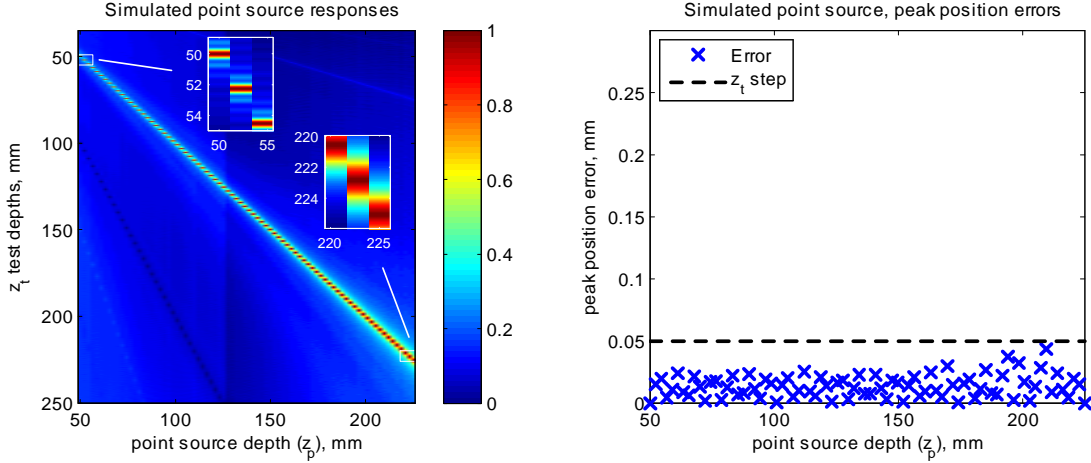


Figure 4-23: $M_c + M_s$ for a point source. Left: a point source is scanned between 50 and 225 mm and the MPFF used to estimate the position. The response is plotted in each vertical line. The width of the peak broadens as z_p increases. The insets depict responses for small and large z_p . Right: the absolute error in the peak position as a function of z_p . The black line marks the step size in z_t . The depth estimate, based on the peak, is within 0.05 mm (the step in z_t).

$N(f_r, f_\theta)$, and the $p = 2$ norm. The f_θ integral is

$$\begin{aligned}
 & \int_0^{2\pi} \left| A(f_r, f_\theta) \cos\left(\frac{2\pi}{\lambda} z_p f_r\right) + N(f_r, f_\theta) \right|^2 df_\theta \\
 = & \int_0^{2\pi} A(f_r, f_\theta) A^*(f_r, f_\theta) \cos^2\left(\frac{2\pi}{\lambda} z_p f_r\right) df_\theta + \int_0^{2\pi} N(f_r, f_\theta) N^*(f_r, f_\theta) df_\theta \\
 & + \int_0^{2\pi} A(f_r, f_\theta) N^*(f_r, f_\theta) \cos\left(\frac{2\pi}{\lambda} z_p f_r\right) \\
 & + \int_0^{2\pi} A^*(f_r, f_\theta) N(f_r, f_\theta) \cos\left(\frac{2\pi}{\lambda} z_p f_r\right) df_\theta, \tag{4.50}
 \end{aligned}$$

$$\simeq \cos^2\left(\frac{2\pi}{\lambda} z_p f_r\right) \int_0^{2\pi} |A(f_r, f_\theta)|^2 df_\theta + \int_0^{2\pi} |N(f_r, f_\theta)|^2 df_\theta, \tag{4.51}$$

if N and A are statistically independent so that $\langle AN^* \rangle = 0$ and the two integrals with cross

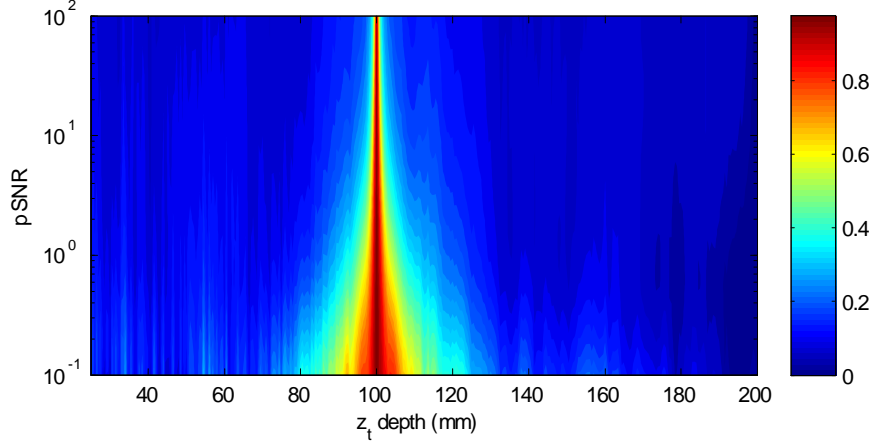


Figure 4-24: $M_c + M_s$ response for a hologram with Gaussian white noise. The object is a synthetic copepod (the fifth object from Figure 4-22) located at 100 mm. The responses for different SNR are normalized for comparison. A strong peak around 100 mm exists throughout the SNR range, with low amounts of added noise.

terms disappear. The spectral estimate is approximately

$$\hat{I}_\theta(f_r) \simeq \left[\frac{\sqrt{2}}{2} \int_0^{2\pi} |A(f_r, f_\theta)|^2 df_\theta + \int_0^{2\pi} |N(f_r, f_\theta)|^2 df_\theta \right] * S(f_r), \quad (4.52)$$

where $S(f_r)$ is a spectral smoothing function. The noise spectrum leads to a bias in \hat{I}_θ and a small amount of spectral weighting in the depth estimate, but less so than the original PFF. Fortunately, the PFF itself was surprisingly resilient to noise, with white noise providing an offset in the F_c estimate and the integration reducing errors from individual spectral points. The MPFF responses with Gaussian noise added to the hologram is shown in Figure 4-24 for the synthetic copepod hologram (the fifth object of Figure 4-22), with the total response normalized to unit amplitude for comparison. A strong peak appears at the correct location, $z_p = 100$ mm, throughout the range of SNRs. The width of the response increases with the noise power, but is incredibly robust to Gaussian noise.

The effectiveness of the MPFF was tested on the set of 174 hand-marked objects in holograms captured by the prototype unit. Plots of example MPFF responses are shown in Figure 4-25 for four randomly chosen objects (the ℓ_1 and ℓ_2 norms are both plotted for comparison). Peaks in the responses are visually easy to find for a few objects, while peaks

p -norm	z_p err, nearest	z_p err, max	Obj. rank	Peak count	Obj. Rank
0.25	2.5 mm	132 mm	6.4	31.7	20%
0.5	2.3 mm	67 mm	5.6	29.6	19%
0.75	2.2 mm	54 mm	5.0	28.2	18%
1.0	2.2 mm	50 mm	4.5	27.4	17%
1.25	2.2 mm	47 mm	4.6	27.0	17%
1.50	2.2 mm	50 mm	4.7	27.2	17%
1.75	2.1 mm	58 mm	5.1	27.4	19%
2.0	2.2 mm	62 mm	5.4	27.6	20%
3.0	2.2 mm	71 mm	5.6	29.5	19%
4.0	2.3 mm	84 mm	6.8	30.9	22%

Table 4.3: MPFF results for estimating the depth of objects in real-world hologram samples. The expected depth error between the known object and the nearest peak in the MPFF response is around 2-3 mm. However, the expected distance from the object to the strongest peak is much larger, around 50-70 mm in most cases, simply due to the fact that the object does not correspond to the strongest peak response in enough test cases. However, the object is expected to be round at the 4th through 7th ranked peak (out of around 30 peaks), 17-20 percent of all the peaks.

for other objects can easily become hidden in the cross-spectral noise. Objects also do not necessarily align perfectly with the peak response or even the largest raw response values. The same multi-scale peak detection algorithm as earlier was used to identify depth regions with a positive response at multiple scales and the peaks ranked based on the maximum response across scales. The known object position was compared against the nearest peak and the peak with the best rank. The position of the object's peak within the ranked list was also recorded. Table 4.3 summarizes the results for different p -norms, reporting the expected values for the depth errors, the position in the ranked peak list that the object's peak was expected to be found, the expected number of peaks per hologram, and the expected percentage of peaks that must be searched in the ranked list before finding the object's depth. In general, the objects of interest were expected to be found within 2 – 3 mm of their true depth and within the first 20% of peaks found by the algorithm. The best results are for $p = 1.0$ to 1.5, with lower numbers of cross-talk peaks detected. These p -norms lie between the blue and green curves of Figure 4-25 and are seen to be a compromise between lower algorithm noise at small p and greater peak contrast at high p .

The ability of the MPFF to detect objects is underrepresented by the expected values

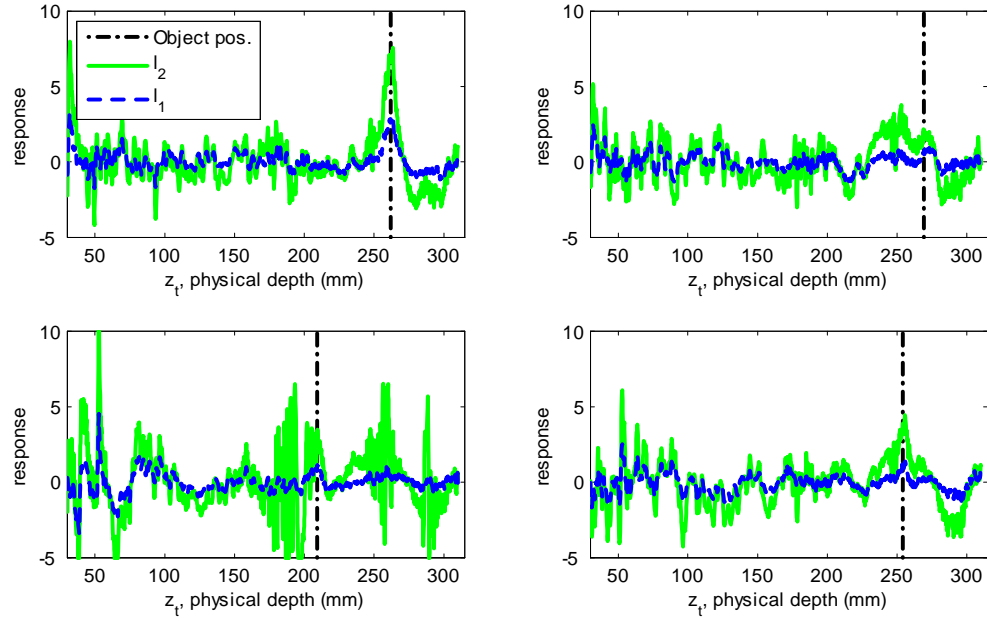


Figure 4-25: Example responses for four randomly selected objects in the test dataset. The true object depth is at the black dotted line; the blue solid line is the $\ell_1 F_c F_s$ response and the dotted green line is the ℓ_2 response. The responses have been normalized against their local means, and the ℓ_1 line had a 10-point smoothing filter applied to show the shape of the curve beneath the noise.

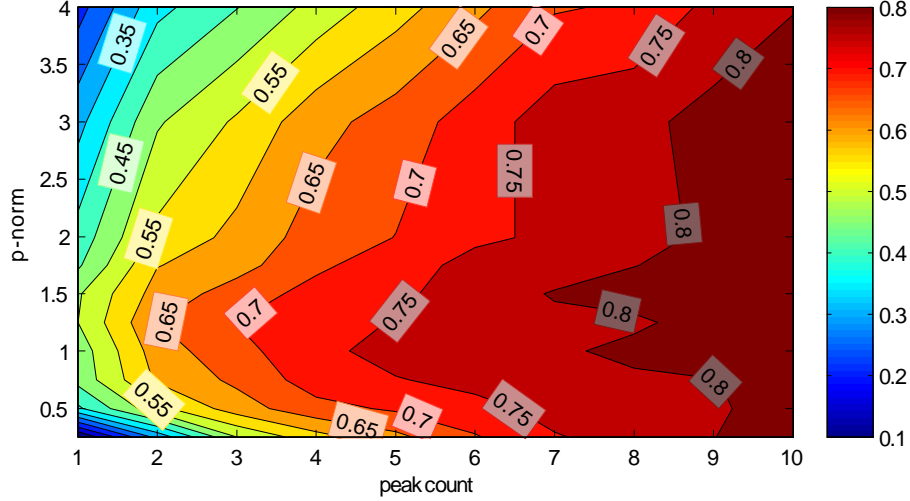


Figure 4-26: Cumulative probability of finding the object at the correct depth as the number of peaks in the ranked list is searched. The best performance comes from $p = 1$ to $p = 1.5$, finding around 70% of objects within the first three peak positions searched.

of Table 4.3, weighted by the objects which disappear into the clutter or are not detected by the particular peak picking algorithm. Figure 4-26 plots the cumulative probability of finding an object within the first n ranked peaks searched. For example, at $p = 1.25$, around 70% of the objects are located at depths corresponding to one of the first three peaks in the MPFF response.

Determining why some objects fail to be detected within the first few ranked MPFF peaks is particularly challenging. The object size, shape, depth, and edge frequency content do not appear to have an influence. Similarly, the total number of objects in the hologram, number of objects within the window used to compute the spectrum, and the hologram noise cannot be statistically correlated to the peak number at which the object is located. Another hypothesis was that substantial frequency content from objects just outside the window was influencing the estimation. To check this, the MPFF was computed for four windows centered to the side of the known objects (but containing the object), and the responses averaged. The probability of finding the object at the first peak increased by about 2%, with the cumulative probabilities for later peaks increasing by 1-3%. This indicates that fringes leaking into the window from nearby objects did not significantly affect the estimation. Unfortunately, checking the statistical independence of the object spectra with

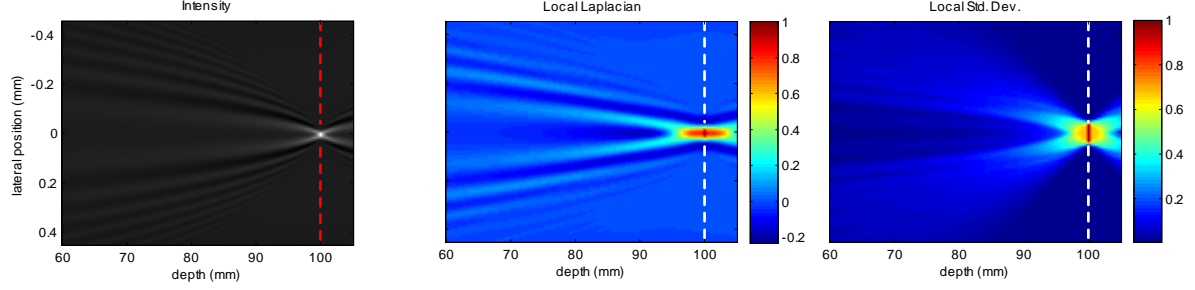


Figure 4-27: Depth response of a local Laplacian and local standard deviation for a point object. The left panel plots the intensity as a function of depth and one spatial coordinate as the point comes into focus at $z_p = 100$ mm (the dotted line). The local Laplacian (center panel) and local standard deviation (right panel), both computed with 9-point windows, have responses that extend into the depth direction. Note that the Laplacian has negative values in its kernel, resulting in negative focus metric values far from the point object.

others in the hologram, including particles, is not possible. At this point, the question is left for future research if there is pressing need for significant speed improvements and fewer reconstructions.

Using the results from the MPFF is a simple matter of reconstructing an image and applying a focus metric to determine whether there is an object in focus at a particular depth. If not, an image is reconstructed at the depth of the next ranked peak until an object is found. The MPFF has an expected depth prediction error of around 2 mm (Table 4.3), so that a focus metric with an extended depth response would limit the number of reconstructions necessary to determine if an object is in focus at a peak location. Of the six basic metrics introduced in Sections 4.1.1 and 4.1.2 (see specifically Table 4.1), the local Laplacian and local standard deviation (square root of local variance) have the greatest extent in depth for a point source (Figures 4-27 and 4-28). The extended depth response of the Laplacian explains why the coarse-to-fine focusing method proposed by Domínguez-Caballero works well [95].

4.2 Accelerated computations on graphics processing units

Digital holography has always been tied to computational power as a numerical technique. Faster computers in the last decade have brought DH from the realm of academic curiosity to an applicable tool, and with it a demand for larger volumes, finer depth resolution, and

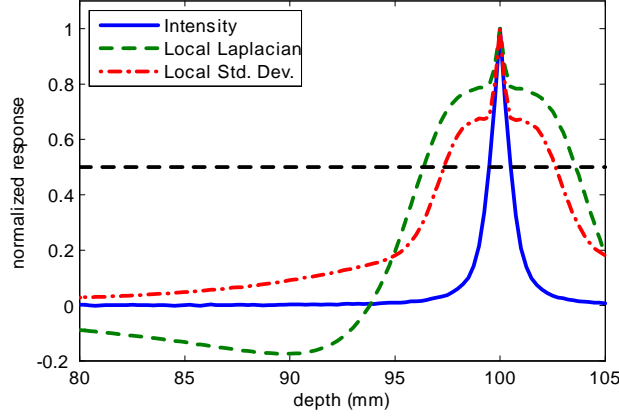


Figure 4-28: Plots of the local Laplacian and local standard deviation focus metric responses as a point object passes through a focus at $z_p = 100$ mm. The maximum response has been normalized to unity, and the horizontal dotted line marks 50% of this maximum. Both metrics decrease slowly with depth compared to the other individual metrics.

larger datasets.

Even with improved CPUs, reconstructions can be time consuming. Consider the example of reconstructing one second of 500 fps video of tracer particles in a fluid flow. A 10 cm distance reconstructed at 0.2 mm depth intervals takes 250,000 reconstructions alone – before applying any focus metrics or analysis to determine particle positions. A single computer which can compute a 2D FFT of a 1024×1024 point array at 15 FFTs per second would require more than 4.5 hours for the reconstruction task. Similarly, reconstructing the entire field for a 16 MPx plankton hologram over a range of 200 mm at 0.2mm depth intervals on a computer which can calculate a 4096×4096 point 2D FFT in 1.0 s would require around 30 minutes. Tracer particles are just one example of dense object fields which require brute-force reconstructions of every plane. Other practical examples include bubble, oil, and other mixed phase flows [380], high plankton concentrations, and holograms with high enough noise that the fast estimation methods fail.

Application of focus metrics also increases the number of calculations. For example, the local Zernike metric of Section 4.1.3 filters the reconstructed image with five kernels. Each filter takes the same amount of time as a reconstruction when the convolution is done using Fourier transforms, so that filtering is actually the most demanding component of applying that focus-detection method.

The fast position and depth estimation methods can reduce the overall reconstruction time by limiting the size of the field and the number of depths which need to be scanned. Even so, reconstructing even a few frames for candidate objects or searching deep within the candidate list (see, e.g., Figure 4-18 and the candidate TP rate versus sorted order for real-world holograms) can still require a significant number of reconstructions.

Holography is well suited to parallel processing since each reconstruction is independent. In cases where multiple holograms can be processed separately, distributing the data across nodes in a supercomputer cluster is reasonable if the network latency is low compared to the processing time [262],[131]. Clusters are a poor solution for a number of other cases. Real-time applications, such as controlling flows based on observed mixed-phase characteristics, require fast turnarounds for processing single holograms; the network latency in distributing the same hologram to multiple processors is extremely high compared to the computation time. Analyzing holograms during a cruise is virtually impossible unless portable clusters become an affordable reality. For that matter, maintenance, hiring programmers, and continued access to clusters can be expensive for labs.

The approach used throughout this work for both the holography of Chapters 3 and 4 and the image processing in Chapter 2 has been to use graphics processing units (GPUs) for computationally expensive tasks⁹. NVIDIA introduced a common Compute Unified Device Architecture (CUDA) for their GPU chips that allows programmers direct access to the underlying parallel processing units for performing general purpose numerical computations. A subset of the C programming language, along with GPU-specific commands, was made available as a software development kit in early 2007. A large list of projects using CUDA for general purpose scientific work has appeared since that time, including a number of projects in optics [229],[331],[332].

The GPU is arranged around multiprocessing units (MPU) (Figure 4-29). Each MPU includes a series of parallel processing cores (there are eight computation cores per MPU in the NVIDIA 9800 GTX, for example, for a total of 128 cores; 112 are available for computation in the 16 MPUs while the remainder are used for controlling execution), a

⁹In fact, the only sections which did not use a GPU within the processing chain were the longline monitoring of Section 2.3, one of the older projects in this thesis, and the fast depth estimation of 4.1.5, which could see benefits from GPU implementation but has not been programmed yet.

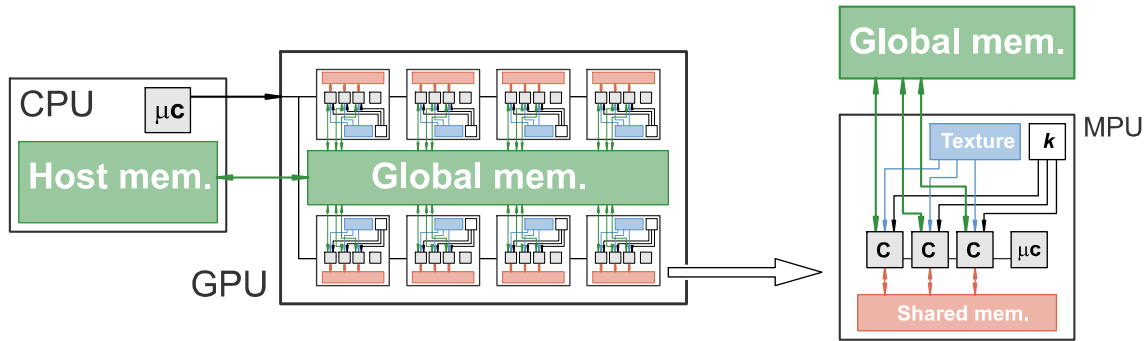


Figure 4-29: Layout of NVIDIA GPUs. Left: the host CPU transfers data between its own memory and the GPU and issues commands. The GPU is a series of MPUs. Right: Information flow in an MPU. C: processor core; μ c: microcontroller; k : constant cache. The cores have read access to each memory, but can only write to the MPU's shared memory and global memory. The shared memory is not exposed directly to the other memories and must be loaded by the processor cores. Three cores and eight MPUs are depicted.

shared memory (16 KB for Compute Levels 1.2 and 1.3, 32 KB for Compute Level 2.0), texture and constant caches (64 KB), and an instruction manager. Each core has its own private memory. The shared memory is readable and writable by the code “threads” running on the multiprocessor, but must be actively loaded with data. The texture and constant caches are mapped by the GPU and are read-only. Reads from the texture cache include special operations to interpolate data on-the-fly in 1D, 2D, or 4D. The instruction manager handles cueing and memory requests for the MPU. A large global memory (512 MB standard for the 9800 GTX, several GB for later cards) can be read and written by the MPUs. The host CPU reads and writes to the GPU global memory and issues commands to the GPU card.

CUDA is designed around single instruction multiple thread (SIMT¹⁰) operations. In SIMT, a small code fragment known as a thread (a “shader” in computer graphics) is run in parallel on the MPU cores. Threads are intended to operate simultaneously in smaller groups known as “warps” (32 threads), running the same commands at the same time. Threads have read-write access to the global memory, MPU shared memory, and the local

¹⁰SIMT is NVIDIA's equivalent of single instruction multiple data (SIMD) parallelism. The difference is that SIMD assumes a certain amount of independence between processors, while SIMT allows collaboration between threads on the same processor. The threads in a single warp are expected to perform simultaneously and are perhaps closer to the SIMD model.

core memory. Interpolation from the texture cache is done by thread-based commands. Math operations were originally designed for 32-bit registers (floats and integers) with fast approximations for certain functions. Since the original release, 64-bit doubles have been added to CUDA, though doubles are generally slower to compute than floats.

Libraries for performing specific operations have proliferated recently. Frigo and Johnson's FFTW library for fast Fourier transforms (the same library used by Matlab), was one of the early ports, and provides support for 1D, 2D, and 3D FFTs (available in the CUDA SDK). Also of interest for optics are the various components of the BLAS (CUDA SDK) and LAPACK (CULA) projects that have also been ported.

The quirks of CUDA determine which algorithms will be better suited for GPUs than others. Memory transfers, access times for global and shared memories, thread ordering, and thread branching can all have a significant impact.

Host-device memory transfers are one of the major bottlenecks. Reading and writing between the host CPU and GPU can be one of the slowest procedures (around 1000 MB/s for a 9800 GTX), so that transferring the data once and processing as much as possible is highly preferable. (In contrast, memory copies within the GPU's global memory can be extremely fast: 42,000 MB/s on a 9800 GTX). For example, take the focus metrics of Section 4.1. The best implementation is to directly reconstruct the slices and apply a focus metric within the GPU so that only the original hologram, the resulting focus measure, and depth map need to be transferred between the GPU and CPU. This can be especially important for small arrays (512×512 pixels or smaller), where the transfer time and latency can be longer than a pair of GPU-based FFTs. When used with the oil droplets of Section 4.3, this approach resulted in a $150\times$ speedup on an NVIDIA GTX 285 compared to CPU processing. (The high speed-up is a strong reason for using projected focus metrics in Section 4.1.) Figure 4-30 plots just the reconstruction speedup alone, where the speedup is compared between an NVIDIA 9800 GTX and a 3.2 GHz Pentium D CPU.

Shared and global device memories each have their own benefits and drawbacks. Accessing global memory is extremely slow, around 400 clock cycles, but can be hidden by the MPU instruction manager: one warp can be set aside while waiting for data, allowing another warp to run. A large number of executing threads helps the GPU hide this latency.

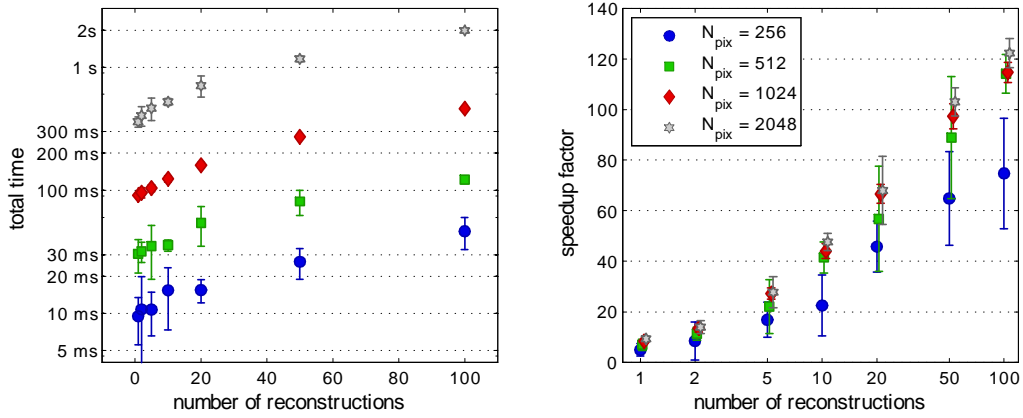


Figure 4-30: Reconstruction component speedups on an NVIDIA 9800 GTX. Left: the total time to compute the requested number of reconstructions (1, 2, 5, 10, 20, 50, and 100). Even for a 2048×2048 pixel hologram, the time to reconstruct 100 slices is manageable at around 2.0 s. Right: the speedup for the same reconstruction operations, compared against a 3.2 GHz Pentium D CPU. Transfer latency and initialization of the GPU FFT structures reduces the efficiency for small holograms. The speedup is especially significant as the number of reconstructions grows and the computation dominates over memory transfers. The speedup for even a single reconstruction ranges from 5 to 9 (256 to 2048 pixels). A $\pm 1\sigma$ uncertainty is plotted for the timing and speedup data. The data in the speedup plot have been shifted laterally by a small fraction for visibility. Note that the timing and speedup has been calculated for the reconstruction step, a single Fourier-domain filter. A series of m additional Fourier-domain filters applied to the magnitude of the field would require $1 + m$ FFTs, making the total time longer by approximately that factor. For example, applying one filter to each of the 100 slices of a 2048 pixel field would take around 6 seconds on this GPU; two filters would take around 8 seconds.

Global memory access is also benefitted by a sequence of threads requesting a sequence of memory, in the same order and without overlaps. This is known as “coalesced” memory reads and writes. Coalescing is a major goal of a well-written thread and a driving reason to design threads that operate on single pixels. Shared memory is much faster, on the order of tens of cycles, but is limited in size and must be pre-loaded by using threads to transfer data from global memory. Small tables whose data is used by multiple threads are ideal for shared memory. Two examples are direct convolutions with small kernels (for example, the local Zernike kernels of Section 4.1.3) or boosted classifiers. Some operations can also be rewritten to efficiently use shared memory instead of global memory. One example used in the CUDA literature is of performing a matrix multiplication by blocks in shared memory instead of columns and rows pulled from global memory.

Discussion of boosting on GPUs deserves some attention given the frequency of its use in this work. The initial implementation of boosting first transferred classifier thresholds and coefficients (Equation 4.20) to shared memory. The threads, one for each pixel, then stepped through each round of the classifier, retrieving the appropriate feature for the thread’s pixel from global memory and comparing its value to the classifier threshold [141]. Decision stumps are preferred in this implementation for limiting the number of global memory fetches, though the approach is amenable to higher-dimensional linear classifiers. Classifiers larger than the shared memory can be used by dividing the classifier into smaller classifiers and combining the results.

An improved implementation for boosting with decision stumps uses texture memory. The stump classifier of Equation 4.20 is first re-factored into a set of look-up tables,

$$Fx = \sum_i b_i + \sum_m A_m(f_m), \quad (4.53)$$

where there are m features, f_m is the value of the m^{th} feature, and

$$A_m(\Theta) = \sum_{i \in \{M\}} a_i(\Theta > \theta_i), \quad (4.54)$$

is the cumulative function for the boosting rounds that use the m^{th} feature in the threshold

test. Specifically, $A_m(\Theta)$ is a step-wise constant function whose value at Θ is the cumulative sum of all a_i corresponding to feature m for decision stumps whose threshold is less than Θ . The A_m functions are uploaded to the GPU and mapped to 1D texture memories. The thread for each pixel steps through the features, retrieving $A_m(f_m)$ using a nearest-neighbor texture fetch. Alternately, the A_m functions can be mapped to a 2D texture memory with each A_m as a column in the texture; the 2D texture fetch aligns with m in the first direction and f_m in the second. Texture caches reside on the MPUs and are managed by the GPU, significantly faster than uploading a full classifier to the shared memory. A decision stump classifier is preferred for texture mapping; higher-dimensional linear classifiers would require an A_m for each combination of features, and is limited to at most a 4D texture.

The SIMT architecture expects the threads in a warp to run the same operations simultaneously. Instructions which branch can cause a warp to diverge if the threads take different paths. When that happens, the threads not on the executing branch are set aside and run separately. Algorithms which limit conditional execution and branching have better execution times. For example, tree classifiers with multiple nodes¹¹ are bad choices for boosted weak classifiers if the classifier is to be evaluated on a GPU. Computing the phase or magnitude of a reconstruction, on the other hand, applies the same operation to every pixel in the image without need for branching¹².

A number of the other algorithms used in this work have recently been ported to GPUs. Examples include SIFT/SURF and image alignments (Section 2.1.5) [77],[194], morphological operations, geometric transforms and coordinate system transforms (Sections 2.1.5 and 4.1.5), histograms (Section 2.1 and pre-processing holograms in Chapter 4) (Section 2.1, direct convolution (Sections 2.2 and 4.1) (CUDA SDK), k -nearest neighbors (Section 2.1) [135], pyramids (Section 4.1.4) (fetches from texture memory), and connected components (Sections 2.2, 2.3.1, and 4.3) (GpuCV). These ports make it feasible to implement the methods developed here on GPUs for significantly faster results and larger data sets, enabling large-scale science at a fraction of the cost.

¹¹A decision stump is technically a tree classifier with a single node and two leaves.

¹²In practice, there is usually some minimal branching: the warps are quantized, so that if the number of total threads exceeds the image size a conditional `if-then` keeps the threads outside the image from reading erroneous memory. The conditional branches only for warps that cross the boundary, thus limiting the number of total branches.

4.3 Application: Oil droplet size distributions

The methods developed and examined in this work have applicability beyond plankton to nearly any microscopic object in the ocean, provided that it has a unique enough visual appearance. One example is low-concentration mixed-phase flows, such as oil droplets or air bubbles in seawater. These have a nearly spherical appearance¹³ across a wide range of diameters, making them a good target for automated methods.

In April 2010, an accident on the *Deepwater Horizon* drilling platform in the Gulf of Mexico began releasing oil and natural gas from a drill site some 4500 feet below the surface directly into the surrounding water at extremely high rates [261]. While some of this oil reached the surface, an unknown fraction stayed in the water column as small droplets – and even then, the question of whether these droplets would surface and where was an unanswered question. Ecologists and modelers did not know the size, depth, or concentration of the drops, all critical for determining the lasting effects of the spill.

Existing instrumentation at the time was extremely limited in its ability to retrieve droplet parameters. Careful laboratory chemical analysis of water collected from depth can determine concentration, but not the droplet size. The size, in particular, helps determine the buoyancy and the rise rate, along with estimates of which ecosystems would be affected. Devices for estimating the particle size distributions (see, e.g., [320]) could not distinguish between oceanic particles and oil droplets and were, in general, limited to shallow depths of only a few hundred meters. Visual inspection from cameras on remote operated vehicles could only indicate the depths at which large droplets could be found. Other imaginative techniques were suggested, such as examining sonar returns [322],[439], but had never been tested against oil droplets.

Holography was an ideal solution to the *in situ* oil droplet measurement problem. Visual measurements can distinguish between oil droplets and other biological objects while simultaneously providing size and shape measurements. A large sampling volume was necessary

¹³Exactly spherical *biological* shapes have been rare in the holograms examined thus far. Two notable examples are fish eggs and diatoms. Fish eggs, however, are nearly perfectly phase matched to the surrounding water and appear transparent; the developing larva is nearly always visible. Laboratory experiments with small spherical and pennate diatoms shows that they appear as point-like or indistinctly small particles in holograms and cannot be distinctly identified as diatoms.

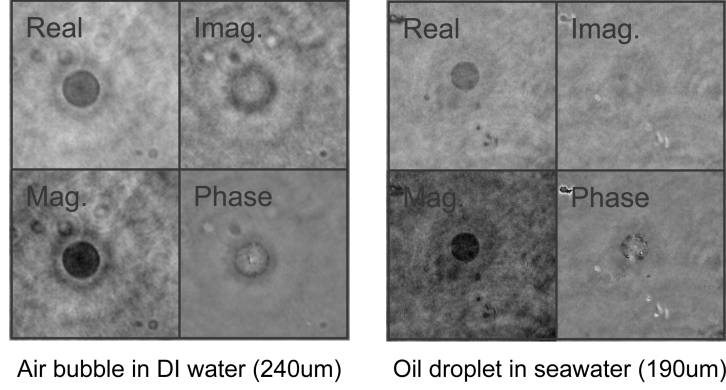


Figure 4-31: Comparison between an air bubble in DI water and an oil droplet in seawater; each quadrant is a different component. Phase is scaled to map over $[0, 2\pi]$. Both the bubble and the droplet appear as round, dark objects, despite the fact that the bubble is transparent and the oil is semi-transparent. The diameter of each is noted in parentheses.

to find and adequately sample the oil distributions. This section describes how those measurements were made with the prototype digital holography unit, employing the techniques developed for object detection and focus determination in holograms. New approaches for estimating the size, concentration, and distributions of the droplets are also introduced.

4.3.1 Droplet models

Droplets and bubbles appear to our visual senses as round, transparent objects, usually with little coloration. The same in holograms appear as dark, round objects (see, e.g., Figure 4-31). Understanding this difference gives a predictive mechanism for round objects embedded in a medium with an unmatched refractive index.

Modeling a droplet can be done by combining a thick lens in the center of the drop with an opaque annulus around the outside, Figure 4-32. The thick lens accounts for the phase difference across the droplet and focusing effects. The annulus models two behaviors: total internal refraction (TIR) and large angles, both of which restrict the ray angles and the intensity transmitted through the drop and subsequently recorded by a detector. If the refractive index of the droplet, n_d , is less than the refractive index of the medium, n_m , such as for air in water (bubbles) or low index materials, the rays beyond $r = an_d/n_m$ fail to enter the drop due to TIR (depicted in Figure 4-33); here a is the droplet radius and r is the

radial distance from the optical axis. In the case where $n_d > n_m$, only a small fraction of rays fail to exit the droplet due to TIR conditions for reasonable ranges of n_d and n_m . The majority of the loss then comes from rays bent beyond the optical system's entrance pupil. The ABCD matrix helps determine exactly where this occurs. The thick lens relations [155] give the matrix system

$$\begin{bmatrix} \alpha_{out} n_m \\ r_{out} \end{bmatrix} = \begin{bmatrix} 1 - \frac{n_d - n_m}{n_d} & -\frac{2(n_d - n_m)n_m}{an_d} \\ \frac{2a}{n_d} & 1 - \frac{n_d - n_m}{n_d} \end{bmatrix} \begin{bmatrix} \alpha_{in} n_m \\ r_{in} \end{bmatrix}, \quad (4.55)$$

where α_{in} and α_{out} are the input and exit angles of the rays, and r_{in} and r_{out} are the distances of the input and exit rays from the optical axis. For a plane wave, $\alpha_{in} = 0$ and the relation for the output angle becomes simply

$$\left| \frac{-2(n_d - n_m)n_m}{an_d} r_{in} \right| \leq (n_m \alpha_{out})_{cr}, \quad (4.56)$$

where the term $(n_m \alpha_{out})_{cr}$ denotes the maximum angle in the medium which can later be recorded by the optical system. This angle may be a function of distance, the NA, and whether the end user requires that the signal is not aliased or is prepared to use generalized sampling [356] to process holograms. For oceanic in-line holography, $(n_m \alpha_{out})_{cr}$ is around 2° to 4° for pixel sizes of 10 microns or less, distances of a few millimeters to centimeters, and the assumption of no aliasing. Thus, r_{in} is small. The fraction of light transmitted which passes through the droplet and on to the detector is given by the ratio of the clear area to the total area, r_{in}^2/a^2 , so that the transmitted intensity is much smaller. This explains why both droplets and bubbles have a dark appearance: much of their intensity is lost due to TIR or high angles.

Bubbles and droplets should also exhibit a focusing effect in their holographic reconstruction. The thick lens system, 4.55, directly provides the equivalent focal length as

$$f_{eq} = \frac{an_d}{2(n_d - n_m)n_m}. \quad (4.57)$$

Using $n_d = 1.49$ for crude oil [358] and $n_m = 1.33$ for water, $f_{eq} = 3.5a$. Similarly, for $n_d = 1.0$ and $n_m = 1.33$, such as an air bubble in water, $f_{eq} = -1.14a$. Both cases have

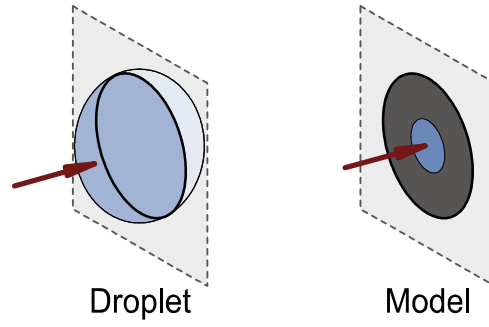


Figure 4-32: Model for an oil droplet as a thick lens (blue inner circle) and an opaque annulus.

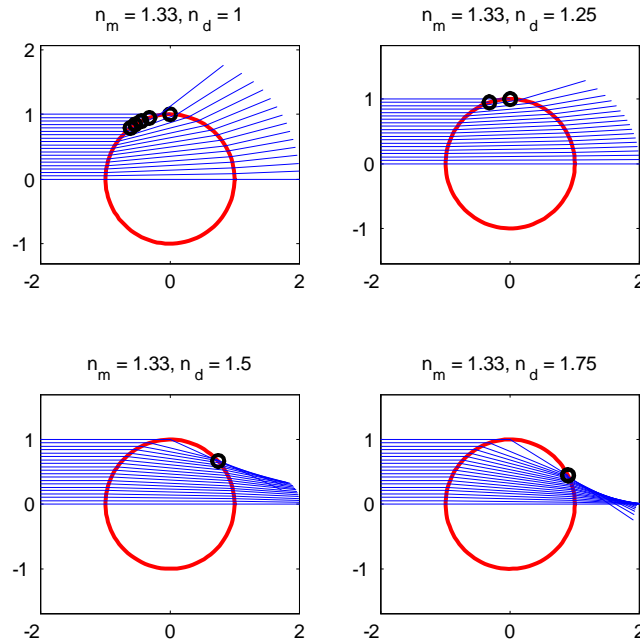


Figure 4-33: Ray trajectories through a spherical volume with index n_d embedded in a medium with index n_m . The red outline is the edge of the sphere, the blue lines are the rays (entering from $\alpha_{in} = 0$), and black circles indicate where rays are lost due to TIR.

focal points which are on the order of the droplet size – and for small drops, on the order of the depth of focus. Even so, the focal effect should have extremely low contrast: for $(n_m \alpha_{out})_{cr} = 3^\circ$, $n_d = 1.49$, and $n_m = 1.33$, the ratio $r_{in}/a = 0.18$ and the transmitted fraction is $r_{in}^2/a^2 = 0.033$.

4.3.2 Detection using focus metrics

A three-step scheme was adopted to locate oil droplets in hologram images using ideas from prior holographic work and from circle fitting work in computer vision. The first step uses a combination focus metric with dense reconstructions on a GPU to locate in-focus edges of objects of unknown shape and size, with the added benefit of improving depth detection in noisy holograms. The second step uses groups of strong edge detections to quickly estimate the circle center and radius, and provides an initial method for discriminating round objects from irregular particulate matter. The third step uses a decision tree with more computationally expensive features to discriminate between oil and particles after the detection has been performed.

Previous studies measured oil droplet sizes ranging from 5 to 500 microns depending on the dispersant, type of oil, ocean turbulence, and depth below the surface [73],[358]. The droplet size distributions tended to follow similar distributions as oceanic particles [330],[256]. Common models used include log-normal, beta, and gamma distributions, which all have a common feature of estimating large numbers of small droplets (see, e.g., the particle size distribution plots of Figure 3-30). Any detection method for oil droplets thus needs to work over a large size range. The recall rate also needs to be high to obtain accurate measurements of concentration and size distributions.

One of the hallmarks of most detection schemes in holography is that they either search for point-like particles or search for unknown objects. Oil droplets provide a luxury of knowing *a priori* the shape of the object. Novel methods of direct detection were first tried on the dataset in an attempt to specifically locate round objects in the holographic images. For each reconstruction plane, circle detection algorithms were applied to locate droplets. Many of these methods were confounded by one of three problems. Hough and Hough-like accumulator methods gave poor results due to the wide range of possible droplet

diameters and center positions, resulting in large parameter spaces and slow computation [438],[276],[307],[389],[237]. Methods which attempt arc fitting are more sensitive to an overall circular shape in the image and lead to a high false positive rate due to twin images and smooth curves from diffraction rings [18],[130],[431],[436]. Finally, methods which use filter-based accumulators, and especially phase accumulators, were sensitive to noise (including multiple nearby detections) and require *a priori* estimates for the minimum and maximum circle size [14],[445]. The end result was that detecting circular shapes directly on the hologram reconstructions was not reliable for oceanic data. However, for constrained systems where the noise can be controlled and the majority of the objects will be circular, such as for laboratory explorations of mixed-phase flows, filter-based accumulators such as the phase-coded annulus are highly recommended [14],[445].

More traditional approaches, volumetric detection schemes using focus metrics, have been reported both here and in previous literature. The most applicable comes from prior work with Tian and Domínguez-Caballero, imaging bubbles of widely varying sizes [380]. Tian et al. elected to use a reconstruct the volume and record the minimum intensity metric [103],[11]. (Note that particle metrics such as de Jong’s complex signature [89] or Pan’s minimum imaginary component variance [274] are inappropriate for these cases since the object cannot be considered to be point-like.) The assumption of minimum intensity is valid when the particles go out of focus quickly and the noise level is low, so that noise does not significantly impact the intensity and the object edges have a low value for only a few reconstruction slices. Particulate matter and large sample volumes for ocean water tend to break the assumptions, leading to gray halos around objects and poor localization—so that the minimum intensity projection is not reliable for determining object size, let alone whether an object is an oil droplet or any number of other marine particles.

Combinations of metrics, Section 4.1.2, were computed on sample oil droplet images. The *SIM* metric, Equation 4.18, was found to be a good edge detector given the noise in the holograms. Estimating the 3D position of an object is done in two steps. The first step is to threshold the *SIM* metric, selecting pixels with a high degree of focus. Those pixels are grouped using morphological dilation and erosion procedures, creating a mask of pixels with a high probability of belonging to an edge. Secondly, the depth is determined by

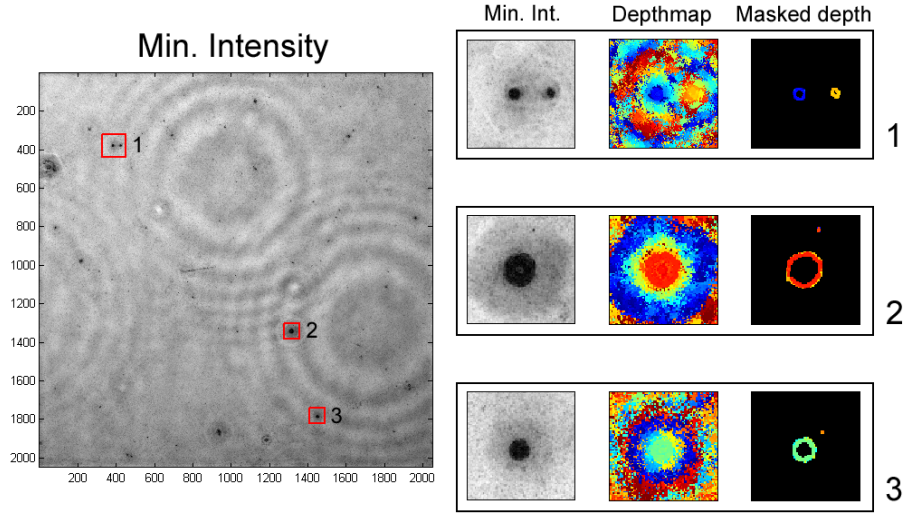


Figure 4-34: Minimum intensity metric applied to a hologram of oil droplets. Zoomed regions of three oil droplets are shown to the right with the minimum intensity, depthmap derived from the slice where the intensity was minimized, and a particle edge mask used to estimate the object depth. The depth ranges from blue (minimum) to red (maximum) following the Matlab *jet* color map.

analyzing the slice at which each pixel's *SIM* value was maximized. In practice, the mean of the depths of the masked pixels is sufficient to estimate object depth. If more precision is required, outliers can be removed by thresholding or by use of mixture models.

Comparisons of using the *SIM* metric and the minimum intensity metric with oceanic oil droplet holograms are shown in Figures 4-34 and 4-35. The minimum intensity metric tends to include noise pixels which can substantially alter the estimated depth, seen as the noisier depth map around the droplet position. The *SIM* metric gives a good prediction of the particle edges, so that direct thresholding or hysteresis thresholding [368],[54] produces a clean edge map. The derived depth map is smooth around the particle edge due to the spatial extent of the filter. The *SIM* results also have high contrast compared against the background: extraneous diffraction rings are not present, unlike the minimum intensity metric.

The *SIM* metric has better localization properties than a minimum intensity metric. The inherent smoothing and intensity weighting remove many of the false positives stem-

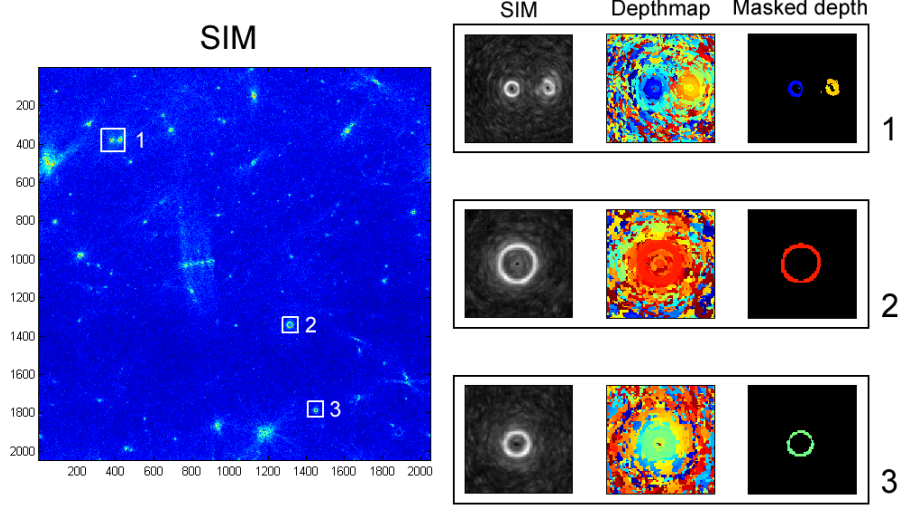


Figure 4-35: *SIM* metric applied to a hologram of oil droplets. Zoomed regions of three oil droplets are shown to the right with the *SIM* value, depthmap derived from the slice where the *SIM* metric was maximized, and a particle edge mask used to estimate the object depth. Depth ranges from blue (minimum) to red (maximum) following the Matlab *jet* color map.

ming from noise, so that the metric is more likely to maximize near a focused edge. Figure 4-36 illustrates the expected error using the *SIM* metric for the case of eighty simulated oil droplets with Gaussian noise added to the hologram. Images were reconstructed every 0.5 mm from 20 to 250 mm. Even at high signal-to-noise ratios (SNR), the *SIM* metric out-performed the minimum intensity metric significantly; for SNRs of 10^{-1} and higher, the expected error was less than 0.9 mm using *SIM*. Depth estimation based on a minimum intensity suffered from noise, leading to outliers and large deviations in the masked depth estimate. Using a Gaussian mixture model (GMM) to remove some of the outlying components improved the estimate by a factor of two for high SNR holograms to around 4 mm of expected error. Note that masking the depths derived from the minimum intensity using the *SIM* metric, the dashed green curve, leads to high errors as more background pixels are selected.

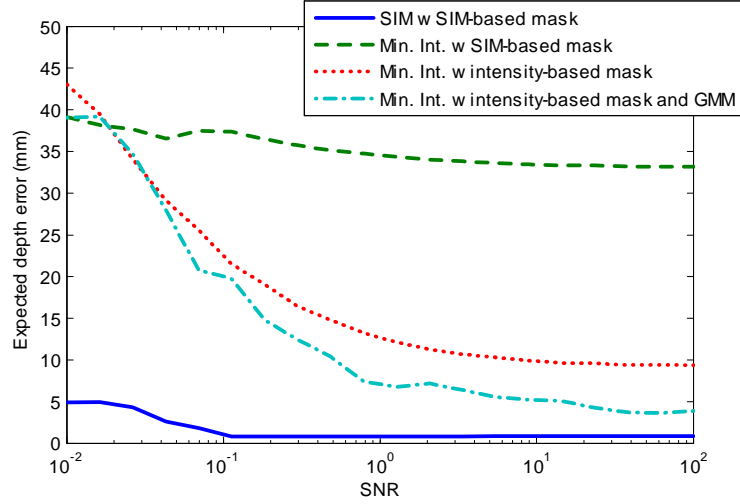


Figure 4-36: Expected depth localization error from SIM and minimum intensity metrics. The SIM-based mask uses edge pixels which have a corresponding SIM value above a fixed threshold. The intensity-based mask similarly uses edge pixels with intensities below a fixed threshold; morphological operations are applied to limit the pixels to just the edge pixels. The final curve includes a GMM to estimate coherent detection clusters, similar to [380].

4.3.3 Fast circle parameter estimation

The first stage located objects in holograms using a focus metric applied to multiple reconstruction slices. The objects detected are indiscriminant aside from having well-defined edges and dark intensities, as predicted by the oil droplet model – but the same characteristics are applicable to many other objects. The second stage seeks to separate out round objects by fitting a circular model to the detected edges. Those objects which have poorly defined centers, centers significantly far from the edges, or with poor fits to the radius parameter can be removed from the search. Objects with well-defined diameters proceed to the next stage of classification. The diameter information is retained for calculating droplet size distributions, volume, and buoyancy characteristics.

Previous methods developed for estimating parameters of circles (center point and radius) tend to fall, again, into three categories consistent with detection methods, as one way of detecting a circle is to determine how well a circle model fits the given data: Hough and Hough-like accumulators [438],[276],[307],[389],[237], direct parameter fitting [130],[431],[436], and random algorithms [62],[67],[295]. As mentioned, the accumulators

require a large parameter space and lack precision. Direct parameter fitting can have varied performance, though it tends to be influenced by noise and deals poorly with small particle sizes. The class of random algorithms tend to be much more robust to noise, especially when the object is non-circular or the detected edge contains a high ratio of outliers to data points.

The random algorithm adopted here makes use of the edge magnitudes and orientations to vote for a center point, then the radius is fit to the edge data. An image of the droplet is reconstructed at the depth detected by the *SIM* metric, and the edge magnitude and orientation are computed using a steerable gradient filter (Figure 4-37). Morphological operations are applied to link nearby detections into a single grouping. The N pixels in the grouping with the highest edge magnitude are then used for estimating the center. First, two random pixels are selected from the N candidate points. The edge normals passing through the points are computed from the orientation information. The intersection point (x_i, y_i) is recorded, along with a weight, w_i . The weight is determined by both the edge magnitudes and the cross product of unit vectors in the direction of each normal as

$$w_i = (S_{M,1}S_{M,2})^{1/2} |\vec{n}_1 \times \vec{n}_2|^2, \quad (4.58)$$

where $S_{M,1}$ and $S_{M,2}$ are the edge magnitudes of the two points and \vec{n}_1 and \vec{n}_2 are the normals. The cross product decreases the weight when the two lines are nearly collinear and are thus more likely to have greater error in locating the intersection point. The circle center, (X_c, Y_c) is then calculated as a weighted average,

$$X_c = \frac{\sum_i x_i w_i}{\sum_i w_i}, \quad (4.59)$$

$$Y_c = \frac{\sum_i y_i w_i}{\sum_i w_i}. \quad (4.60)$$

Once the center has been estimated, the radius is also fit using a weighted average, this time using the edge strength as the weight,

$$R = \frac{\sum_{j=1}^N S_{M,j} \left[(x_j - X_c)^2 + (y_j - Y_c)^2 \right]^{1/2}}{\sum_{j=1}^N S_{M,j}} = \frac{\sum_{j=1}^N S_{M,j} r_j}{\sum_{j=1}^N S_{M,j}}. \quad (4.61)$$

A final fit metric, FM , is computed by summing the weighted squared errors,

$$X_{c,var} = \frac{\sum_i w_i (x_i - X_c)^2}{\sum_i w_i}, \quad (4.62)$$

$$Y_{c,var} = \frac{\sum_i w_i (y_i - Y_c)^2}{\sum_i w_i}, \quad (4.63)$$

$$R_{var} = \frac{\sum_{j=1}^N S_{M,j} (r_j - R)^2}{\sum_{j=1}^N S_{M,j}}, \quad (4.64)$$

as

$$FM = [X_{c,var} + Y_{c,var} + R_{var}]^{1/2}. \quad (4.65)$$

The fit metric is admittedly *ad hoc* in nature as it attempts to combine biased variance estimates for the center position and radius. In practice, the metric is responsive enough to allow preliminary discrimination. The $X_{c,var}$ and $Y_{c,var}$ estimates both grow large if the object points do not agree on a common center, such as for elliptical or rectangular objects, while the R_{var} estimate grows large when the shape is non-circular or includes concentric circles.

The circle parameter random accumulator (CPRA) described here differs from earlier work in a number of ways. First, it does not strictly rely on an assumption of circularity, using only two points with edge normals to vote for a center point of an arc; for example, measuring quadratic curves with K-RANSAC[62] and Cheung’s triplet line method [67] require three points to determine circular arcs, while the gradient vector pair method [295] uses pairs of gradients from opposite sides of the circular edge. Secondly, the CPRA does not require a complete arc like the gradient vector pair method [295] or Barwick’s chord method [18]. A final note is that the CPRA is related to accumulator methods but with a randomly selected subset of high-fidelity edge point pairs (as measured by the gradient magnitude) and a continuous-valued accumulator.

The CPRA requires only a limited number of computations. The filtering operation to generate candidate edge points is common to almost every parameter estimation method and is $O(2P^2 n_{filt}^2)$ when done with direct convolution, where P is the number of pixels per edge of a droplet image and n_{filt} is the number of pixels per edge in the filter. The remainder of the CPRA is $O(n_{iter})$, where n_{iter} is the number of line intersections computed in the

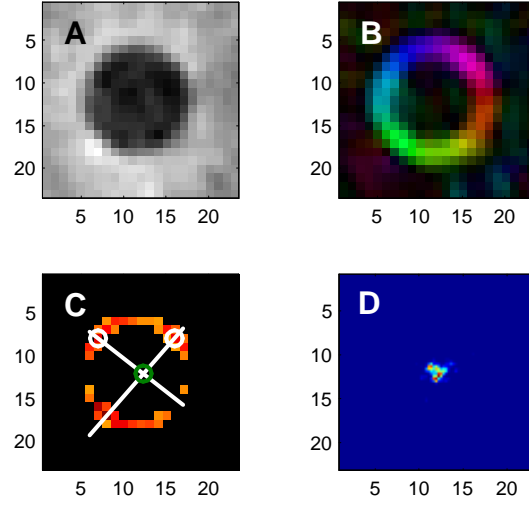


Figure 4-37: Overview of the CPRA method for estimating circle parameters. A, the magnitude image of a reconstructed oil droplet. B, the steerable gradient; hue represents the edge angle and the brightness is the maximum gradient magnitude, S_M . C, the $N = 40$ pixels are used to estimate the center by finding the intersection point (green circle) of lines passing through two randomly selected edge pixels (white circles). This process is repeated n_{iter} times. The weighted distribution of candidate center points is shown in panel D, forming a small cluster near the center. The distribution uses the Matlab *jet* colors, where blue is a low value and red is a high value.

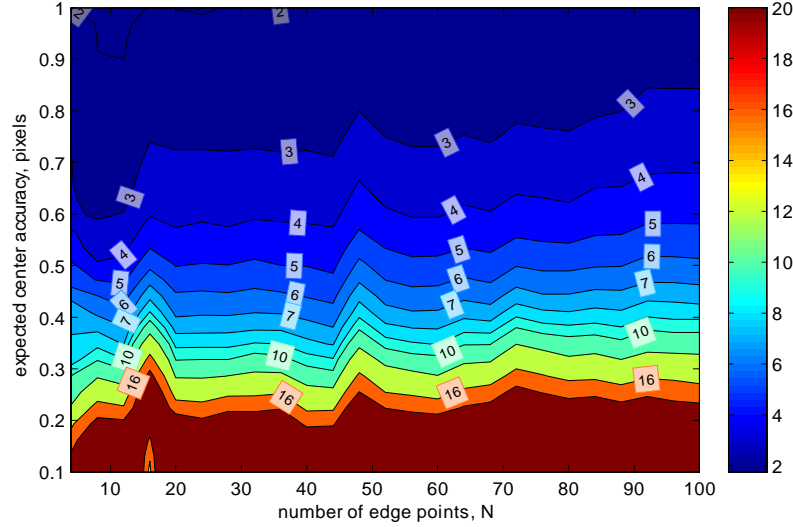


Figure 4-38: Expected number of iterations, n_{iter} , in the CPRA method to achieve a desired accuracy in locating the center point of an example reconstructed oil droplet image. Contours are every integer from two to ten, twelve, sixteen, and twenty.

accumulation process. Note that if N candidate edge points are used, the number of unique intersection points is $N(N-1)$. This gives a natural maximum limit to n_{iter} . In practice, both N and n_{iter} can be surprisingly small, with $n_{iter} \ll N(N-1)$. Figure 4-38 shows the expected n_{iter} to achieve a desired subpixel accuracy in estimating the center point of the circle for the sample oil droplet of Figure 4-37. The center is located to better than 0.4 pixels for $n_{iter} = 10$, on average, and is nearly independent of N . The slight decrease in accuracy with higher N is due to the inclusion of pixels off the edge but with high steerable gradient magnitudes.

A set of simulated droplets was used to estimate the performance of the CPRA and compare it against filter-based phase-coded annulus methods [14],[445]. Holograms of 80 simulated oil droplets were computed and Gaussian noise added with varying SNR. Images of droplets were reconstructed at a depth estimated using the *SIM* focal metric. The magnitude of the field was used to compute the steerable gradient and gradient orientation. The $N = 20$ pixels with the highest gradient magnitude were used with the CPRA to compute the center positions and radii. Phase-coded filters were also applied to the gradient magnitude image with a minimum radius of 1.5 pixels and a maximum radius of half the

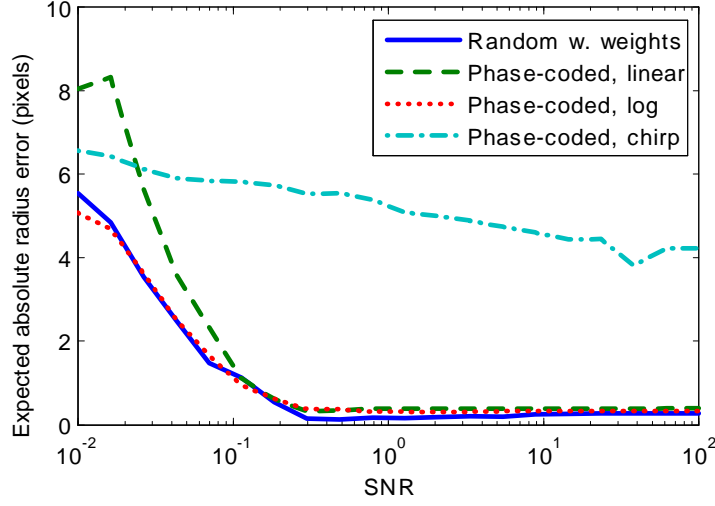


Figure 4-39: Expected absolute error in droplet radius estimation using the random algorithm discussed and three phase-coded annulus schemes. The logarithmic and random algorithms have similar performance over the range of SNR values tested.

largest image dimension (including an additional padding margin around the estimated particle image of 11 pixels on each side). The mean error for each SNR was used to correct the raw data against bias. Figure 4-39 summarizes the resulting expected absolute error in computing the radii. The CPRA, linear-encoded, and log-encoded filters give good results over a wide range of SNR values, while the chirp encoding is much more variable.

Varying the aspect ratio of the simulated droplet also affects the expected accuracy, shown in Figure 4-40. A set of thirty simulated oil droplets were generated using a random diameter distribution ranging from 2 to 52 pixels, then reshaped to aspect ratios between 1.0 (purely circular) to 2.0 (ellipse with major axis twice as long as the minor axis). The deformed droplet models were used to generate holograms and Gaussian noise was added at a fixed SNR of 1.0. The depth detection, reconstruction, and parameter estimation proceeded as previously. The CPRA and linear phase-coded annulus filter have similar expected errors, while the logarithmic encoding shows lower sensitivity, as expected [445]. The chirp encoding has poor results and will not shown or pursued further. The geometric radius of the deformed droplet, $(ab)^{1/2}$, where a is the semi-major length and b is the semi-minor length, is used as the equivalent radius for computing the radial error.

Practical use of the CPRA and phase-coded methods for diameter measurements show

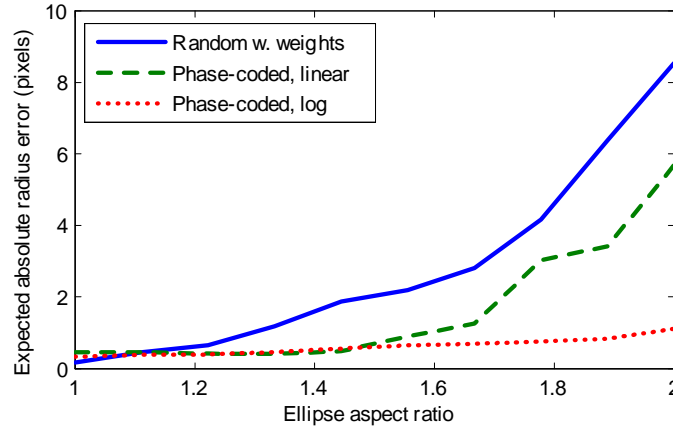


Figure 4-40: Expected absolute error in radius estimation as the aspect ratio of the droplet changes from circular (aspect ratio of 1.0) to twice as long as wide (2.0).

that the CPRA actually performs better than expected (Figure 4-41). Noise, wider varieties of shapes, variations in edge magnitudes, and small droplets seem to affect the phase-coded method more – especially if *a priori* estimates for the diameter range are used (which can be useful if only a portion of the droplet is detected and thus the detection size is a poor initial guess for the object size). A set of 135 oil droplet images was used to calculate linear correlation coefficients between the estimated diameter and diameters measured by hand. The CPRA method resulted in a correlation coefficient of $\tau = 0.73$, while a linear phase-coded annulus had a correlation coefficient of only $\tau = 0.45$. The difference is that the phase-coded annulus overestimated the size of small droplets due to the *a priori* droplet size guess, while under-estimating the size of large droplets. In the end, both estimation methods were used as features in a classification algorithm, with the CPRA diameter used for the final size statistics.

4.3.4 Droplet classification

The final step is to classify images as either *droplet* or *non-drop* and compute the distribution statistics. A training set of droplets and non-droplets, as marked by a human oracle, was used to train a classifier. The set included 118 drops with diameters larger than $30 \mu\text{m}$, large enough to be identified reliably as oil droplets by the observer, and 314 non-drop objects. A series of metrics were computed on the images, including the CPRA and phase-coded

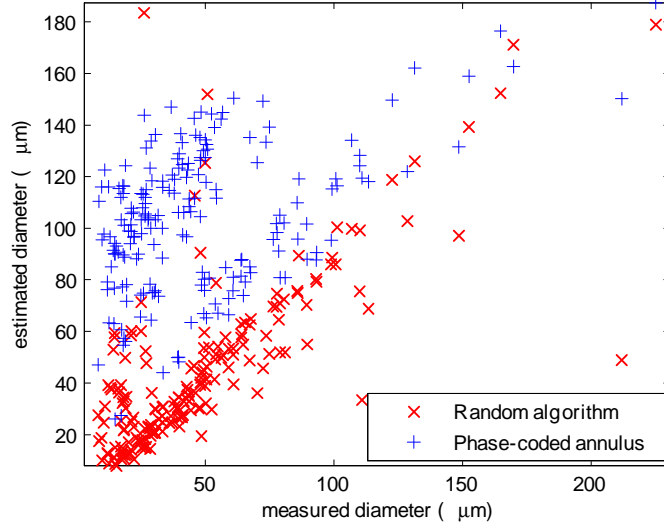


Figure 4-41: Particle diameters as estimated by the CPRA and phase-coded annulus methods. Ground truth was provided by reconstructing and measuring particle diameters by hand for one dive. The phase coding was linear and *a priori* diameter ranges were used due to noise in detecting droplets. The phase-coded annulus tends to strongly over-estimate the size of small droplets when strong noise and weak edges are present.

annulus diameters, average *SIM* value along the edge, measures of circularity computed from the moments of the particle image (see, e.g., discussion in [403], in particular the background on using Hu moments as circularity measures; this is related to the eigenvalue ratio of the intensity Hessian), the strength of the phase-coded annulus maxima, the CPRA fit metric, mean intensity and standard deviations inside the estimated circle area and outside of it, and the magnification factor. Various classifiers were tested in Weka [152],[391] with 10-fold cross-validation on each classifier. A naïve Bayes classifier [106] had an 83.6% overall correct classification rate and provided a baseline. LogitBoosting [129] and J48 trees (the non-commercial version of Quinlan’s C4.5 trees [292]) both resulted in around 87% correct classification, while a J48-Graft tree [416] led to a slightly better 88.7% classification. The J48-Graft confusion matrix is shown in Table 4.4. Interestingly, the J48-Graft can also be used on droplets with diameters smaller than 30 μm . However, the error rate in this case is unknown because the correct label was unknown even to the humans visually labelling the detections.

The classification rates depend on the droplet size. Small objects have higher classifica-

	Est. Drop	Est. non-drop	Truth rate
Drop	91	27	77.1%
Non-drop	22	292	93.0%

Table 4.4: Confusion matrix using a J48-Graft tree for classification

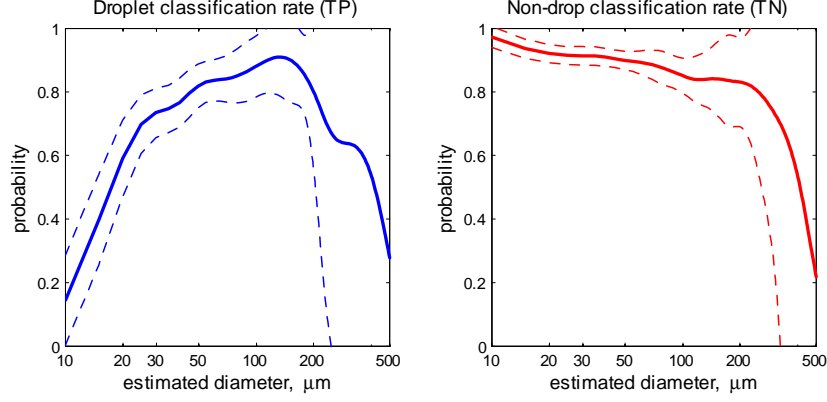


Figure 4-42: Truth rates for droplets and non-droplets as a function of diameter. The rates are estimated from the J48-Graft tree classifier using a kernel density estimator. Dashed lines denote the 2σ uncertainty, derived from the density estimation.

tion errors due to limited image sizes and edge points. Large objects also faced error due to limited statistics, as most objects were $150 \mu\text{m}$ or smaller (equivalent circular diameter). The truth rates as a function of diameter are shown in Figure 4-42. The rates were computed using the training set with the J48-Graft tree and a kernel density estimator (KDE). A logarithmic transform was used in the KDE since the diameter distribution is strictly positive.

The observed number of objects for a particular size class can be modeled as a linear mixture of correctly and incorrectly classified objects. The observed number of objects, $\vec{o} = [o_{drop}, o_{nondrop}]^T$, is then

$$\vec{o} = \mathbf{M} \vec{n}, \quad (4.66)$$

where $n = [n_{drop}, n_{nondrop}]^T$ is the actual number of drops and non-drop objects and \mathbf{M} is the classification rate matrix computed directly from the confusion matrix. The least-squares solution is found as $\vec{n} = \mathbf{M}^{-1} \vec{o}$ [347],[165],[166],[423]. One of the issues is that the direct solution can lead to negative results. Hu solves this problem by setting negative counts to zero and distributing the errors to the other classes depending on their relative

abundances. An improvement is to solve the constrained least-squares problem,

$$\begin{aligned} & \text{minimize } \|\mathbf{M}^{-1}\vec{o} - \vec{n}\|_2 \\ & \text{subject to } n_i \geq 0 \ \forall i, \end{aligned} \tag{4.67}$$

which guarantees that the \vec{n} will be physically feasible.

Sample results using the methods discussed in this section are shown in Figures 4-44 and 4-45. The data comes from the Gulf of Mexico during the British Petroleum (BP) oil spill [261]. The site was located at $28^\circ 44.91' \text{ N}$, $88^\circ 22.0' \text{ W}$, about two kilometers from the failed drill site. The displayed data is from the morning of June 14, 2010, taken during the fifth dive on the Jack Fitz 1 cruise. The prototype holographic camera system with a Hasselblad CFV-39 as the detector was attached to the front of a remotely operated vehicle (ROV) and lowered vertically to nearly 4900 ft, crossing the undersea oil plume twice and collecting 1235 useful holograms before filling the available recording memory. Figure 4-43 shows the depth profile of where data was collected. The detection and classification schemes discussed in this section were used to count the number of droplets. Size statistics were computed by grouping together detections at nearby depths so that each region (shown as a different color in Figure 4-43) had the same number of droplets. Alternately, equally-sized depth regions could be used, or regions with equal numbers of holograms. The method here of dividing the regions based on droplet count has the benefit of providing good statistics for each region while allowing adaptive region sizing around the depth of the oil plume, and was selected for depiction here for better visualization.

The oil plume is visible between about 3500 and 3850 ft (1067 to 1173 m) as a sudden increase in the number of small droplets. The narrow range is consistent with other plume measurements [53],[182] and could be verified visually using a UV light which caused the oil to fluoresce visibly in the ROV camera images. The resultant volume estimates, computed from the diameter estimates, are shown in Figure 4-45. Both the size and volume distributions include 2σ uncertainties, calculated by error propagation. This error includes the uncertainties in the least-squares solution, the kernel density estimation using logarithmic transforms, and the diameter estimation. Larger uncertainties appear for small droplets in

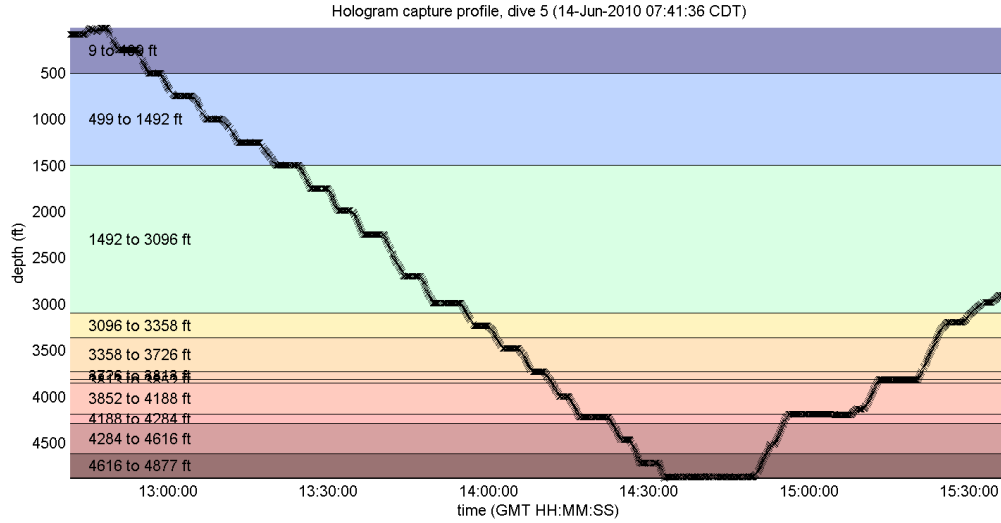


Figure 4-43: Sampling profile of dive 5. Crosses denote depths and times that holograms were captured. Color bands indicate depth regions with equal numbers of detected droplets and correspond to the colors of Figures 4-44 and 4-45. The region size depended on the actual abundance of droplets and the time spent sampling at each depth.

the size estimates of Figure 4-44 because of the lower classification rates for the droplets smaller than $30 \mu\text{m}$. The volume estimates include diameter cubed terms, so that errors in the small droplets count little towards the total volume, but the low numbers of droplets larger than $150 \mu\text{m}$ lead to higher errors due to limited statistics.

In the end, using the holographic camera for investigating the oil spill was successful. Visual measurement was important because the holographic system could resolve droplets much smaller than the ROV's cameras, verifying that the presence of large droplets that could be seen in the UV lights were positive indicators of oil activity. Size distributions that could be strongly linked to oil droplet images were used to improve predictive models for later sampling days, indicating that the modelers should include numerical simulation runs for smaller droplets than previously considered. The distributions could also be used to calibrate sensors which estimated oil content by derived quantities – or in the case of the LISST [320] used on a few later dives, that its size measurements were not sensitive enough to small droplets or discriminative to droplets versus other particulate matter. The oil droplets were a case where visual measurements of small objects distributed through a large volume were critical to the scientific goals.

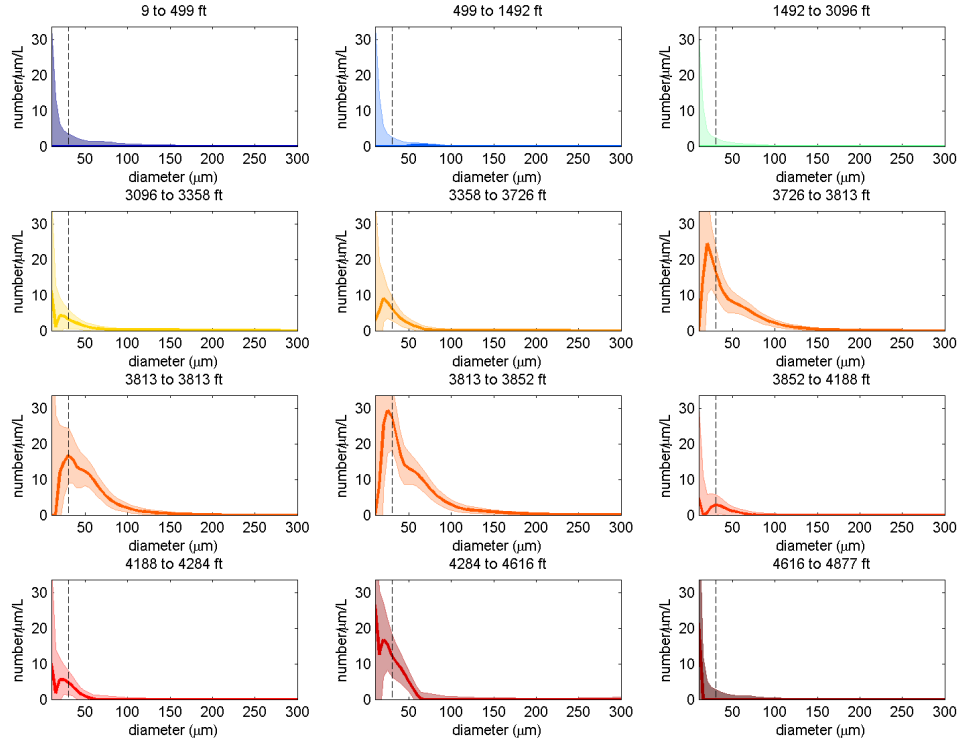


Figure 4-44: Oil droplet size distributions. The solid line is the maximum likelihood estimate; the shaded areas mark the 2σ error as discussed in the text. Regions correspond to those shown in Figure 4-43 and have equal numbers of detected oil droplets for comparative statistics. The dashed vertical line denotes the $30\ \mu\text{m}$ diameter limit below which the classifier error is only roughly estimable. The plume is visible in a narrow depth band between roughly 3500 and 3850 feet with greatly increased probabilities of small ($<100\ \mu\text{m}$) droplets.

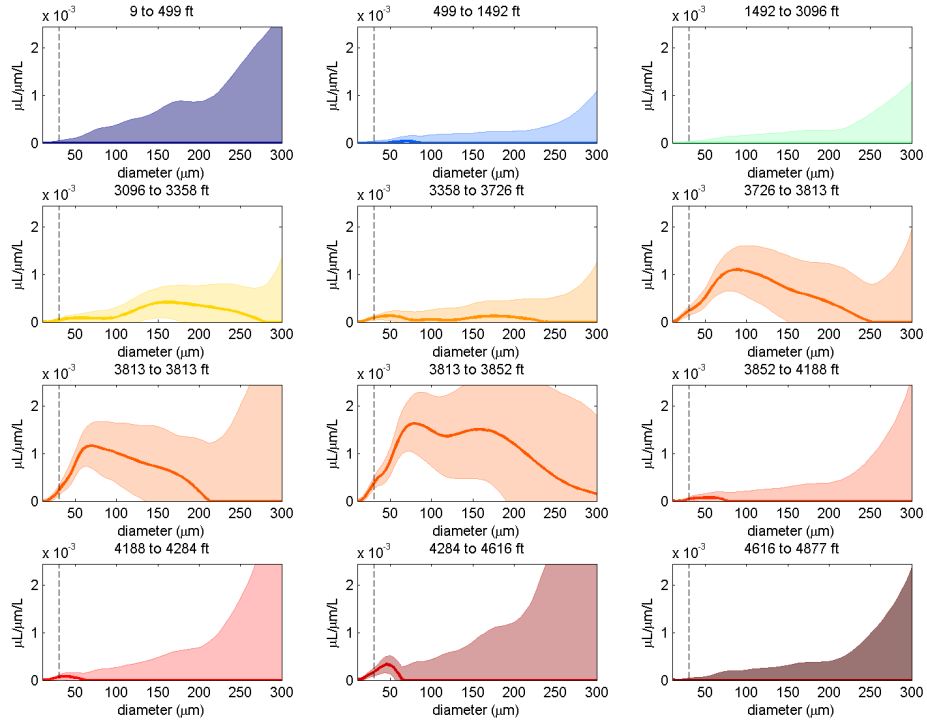


Figure 4-45: Estimated volumetric distributions as a function of the droplet diameter. The solid line is the maximum likelihood estimate, while the filled region denotes a 2σ uncertainty. The regions again match those of Figure 4-43. Extremely limited numbers of large drops were seen in the data, limiting the accuracy at higher diameters.

Chapter 5

Future Work and Conclusions

Tools for making sense of the vast imagery generated by computational imaging inspire a number of further applications. This final chapter serves as a reminder of the contributions that pave the way for future work, then suggests some interesting extensions.

5.1 Contributions

This thesis has applied ideas from machine learning, methods from computer vision, analysis of optical systems, and engineering design to the tasks of optically retrieving information about life in aquatic environments. Solutions have been provided for a number of problems which, at the time, had not been approached or sufficiently solved. Specific contributions of note include the following:

- Chapter 2 developed methods for recognizing habitats and extracting images of fish from images captured using traditional cameras. A bag of words model was shown to be particularly effective with mixed benthic environments with significant improvements achieved through a transform based on an independent component analysis. A multi-stage detector was created for extracting images of rockfish from the same habitat images with an overall 70.9% detection rate and precision of 0.69. Coupling the habitat information with each detection allows the population counts to be corrected from the known detection rates. The detector also provided evidence that the SeaBED AUV is not leading to avoidance behaviors from rockfish. Finally, methods

for extracting images of fish from videos of longline fishing operations are demonstrated, paying particular attention to dealing with a non-ideal background, temporal and spatial resolutions, and high compression rates. A max-flow model was created to predict which extracted regions from nearby frames were of the same fish.

- Wigner transform methods were used to analyze holography with a spherical reference wave, showing an equivalence between spherical and planar holography that simplifies the reconstructions. A linear model was derived for space-bandwidth transfer in holography that uses both a spherical reference and magnification optics, significantly simplifying the relationships between the space and frequency limits of an imaged object and the digital detector. The model is applicable for nearly all in-line digital holography with planar or spherical references. Calculations of the field of view, depth of field, sampling volume at a target resolution, magnification limits, and relationships between the spherical reference parameters and imaging bandwidths are demonstrated using the model. A spherical reference is used in a prototype as a direct result of the predictability of the imaging.
- A prototype *in situ* digital holographic imaging device was engineered in Chapter 3. Theory to understand how motion affects the formation of holographic images and computational experiments to determine how the number of recorded bits affects image quality were used to inform the engineering choices. Filters to reconstruct images recorded with Bayer filters were explored and software to work with high space-bandwidth product holograms was created.
- Theory was developed to explain the noise observed in holograms from particles and matched well with simulations. A ρ^{-a} model for the power spectrum, with ρ as the spatial frequency, was shown to fit experimental holograms, while coefficients of $a = 2$ to $a = 3$ correspond well with all of the common particle models. The theory was extended to spherical reference waves. Computations show that the a coefficient remains essentially the same, independent of the type of reference. Simulations of holograms with particle distributions provide a starting point for predicting image degradation as a function of physical parameters.

- A novel focus metric based on local Zernike moments and boosted classification was created in Section 4.1.3, trained specifically to recognize edges in noisy, real-world holographic reconstructions. The method is applicable to other types of recognition problems, such as particles.
- Two methods for predicting the lateral locations of objects from raw holograms were discussed and a practical approach to identifying peaks suggested. The first method extended the summation kernel to spherical references. The second proposed the use of local frequency estimates to detect areas consistent with a holographic signature. Four approaches to calculating the local frequency were tested. Results for object detection using the summation kernel and a Riesz-based frequency estimate were presented on holograms captured by the prototype unit.
- A fast method for estimating the depth of objects in a hologram based on a normalized spectral signature in the polar domain, the MPFF, was created and tested. Computer-generated holograms showed excellent performance characteristics for the MPFF in its ability to precisely and accurately estimate depth over a range of conditions (depth, noise, shape, etc.). Performance with a particularly challenging set of real-world holograms showed a maximum of around 70% of objects detected within the first three peaks when a p -norm of 1.25 was used.
- The methods developed during the thesis work were applied toward quantitative measurements of oil droplets in a deep water plume. Holograms were captured using the prototype unit, then processing was done on a GPU with an appropriate power filter (Chapter 3). Images of oil droplets were extracted using a novel focus metric, then the size estimated using a new technique that is more robust to the noise observed in real-world reconstructions (Chapter 4). Classifiers with corrections are used to estimate the oil droplet size distributions (influenced by work in Chapter 2). Results of the plume location are consistent with other researchers. The work was complementary as it added size data and direct imaging of the oil as it interacts with the environment that other instruments were not able to provide.

5.2 Applications and extensions

The work in this thesis has focused on providing physical, software, and theoretical tools for imaging life in aqueous environments. As with any set of tools, there are a myriad of applications and uses. This section proposes future research that extends from the methods developed in Chapters 2-4, along with a number of related projects that have shown promise after initial experimentation. The list is, of course, not comprehensive, but gives a flavor for some of the more immediately applicable extensions.

5.2.1 Automatic identification of plankton

The design of the prototype unit was engineering around imaging plankton in the ocean. A large number of holograms have since been captured by the prototype. The methods developed in Chapter 4, specifically localization, focus metrics, and GPU implementation, were one of the final steps before plankton images could be reliably extracted from the holograms automatically. The next step is to select the appropriate focus metrics, tune the parameters, and optimize the code to provide a database of plankton images for training.

Feature extraction and selection for holographic images is not a trivial task. The twin image and noise increase the variability of the background so that simple thresholding provides extremely poor shape outlines. Methods such as Chan-Vese segmentation (see Figure 5-1 and Section 2.2; GPU implementations have recently been made available) or graph cuts [239],[238] have a better chance of correctly determining the plankton shape with fewer priors, particularly important for recognition based on the morphology [334]. Past research on recognition from digital holographic images is extremely limited, suggesting simplistic correlation and statistical methods with limited classes [176],[178],[177],[257],[263]—but does indicate that the imagery is useful for recognition. Fortunately, there is a good selection of more appropriate research showing that plankton identification is indeed possible [29],[165],[166],[234],[81],[370],[147].

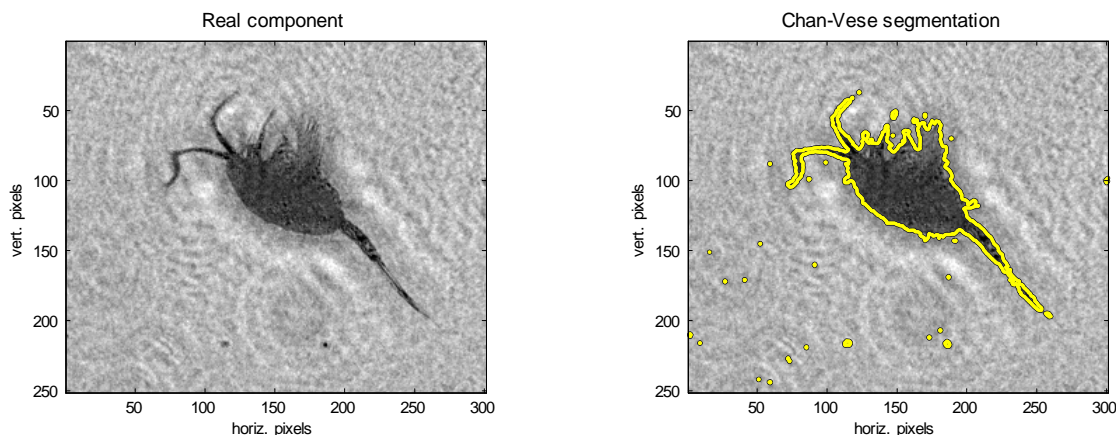


Figure 5-1: Example of using Chan-Vese segmentation with a copepod that has a distinctly different grayscale level than the background. The real part of the field, left, was used to compute the segmentation, right. This segmentation used 100 iterations and took 160 ms to compute for the 0.072 MPx image on a 2.2 GHz CPU.

5.2.2 Particle size distributions

Imaging particles is one of the older and better studied uses of digital holography. Holographic particle imaging velocity tracks particles through 3D volumes to produce estimates of the flow field [160],[273],[290],[240],[328],[441], with a number of specific algorithms developed for dealing with particle localization [274],[89],[435],[83],[325],[348],[11],[13],[336] and pairing between images [327],[432]. Early film holography sought to measure particle size distributions in the environment [377], a trend continued with digital holography [272],[321],[253] (and Section 4.3).

Estimating the particle size distribution is interesting for biological, ecological, and industrial applications. Nishihara et al. proposed a simple method for estimating the size of a *single* particle by searching for the zeros of the spectrum [264], then Denis et al. demonstrated that the mean particle size of a distribution could be found from examining the central spectral peak [92]. However, Section 3.4 showed that estimating the full particle size distribution from the spectrum is extremely challenging and may not be possible without excellent *a priori* information and controls over the measurement. Obtaining a complete particle size distribution is thus relegated to the spatial domain.

Many of the methods in Chapter 4 are applicable to searching for particles in the spatial

domain. Extremely low particle densities could make use of the fast depth estimation and localization methods of Sections 4.1.4 and 4.1.5. Reasonable particle densities, on the other hand (such as those required for holographic particle imaging velocimetry or concentrations seen in oceanic holograms) are better suited to a brute-force reconstruction of the entire volume. Methods which utilize a graphics processing unit, Section 4.2, are arguably necessary for large volumes. The various focus metrics and combinations, Section 4.1, are easily adapted for use on GPUs. An example of using focus metrics on GPUs to locate oil droplets over a range of diameters was discussed in Section 4.3. Other applications require testing and tuning to determine which metrics are best for the particular goal: accurate sizing, lateral localization, or depth localization.

The Zernike method of Section 4.1.3 was specifically trained for edges and gives a weak response for small particles. However, the same concept can be applied to train a classifier for particle images. Initial tests show promise, with the specific caution that noise needs to be controlled to avoid false positives.

The projection focus method is good for GPU implementation (Sections 4.1.1 and 4.2) but removes depth information. This can be a problem when there is noise close to a particle laterally but far away in depth. Removing the noise during the filtering and projection process improves size and position estimation. Initial experiments with non-linear bilateral filters [277] significantly improved the contrast around tracer particles (Figure 5-2) [230]. Metrics which amplify noise, such as gradients and Laplacians, may be particularly good candidates for combination with bilateral filtering. Recent work from Pantelic et al. specifically merges bilateral filtering and edge detection [275], supporting the use of bilateral filters for detection and localization.

5.2.3 Biological tracking

Digital holography has the unique ability to resolve positions of objects within 3D volumes over longer working distances than standard imaging systems. It can thus be used for behavioral studies of biological entities in their natural environments [326],[328],[432],[99]. Figure 5-3 shows an example of using the lateral and depth resolving capabilities of a benchtop system (Section 3.3.2) to record the motion of a copepod over 2.5 minutes as it explores a

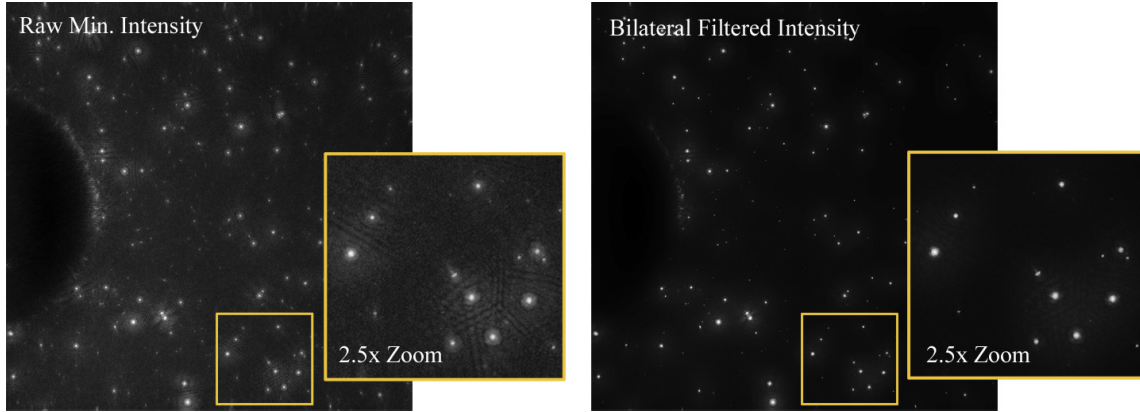


Figure 5-2: Locating particles with a minimum intensity focus metric and bilateral filtering. Left: the projected minimum intensity metric retains noise, both from random background fluctuations and from particle focusing, decreasing the contrast of the particle images. Right: applying a bilateral filter to each reconstructed slice during the projection removes random fluctuations and a significant amount of the out-of-focus image, increasing particle contrast.

small enclosure (pictured as the black boundary box). Details about the orientation of the copepod and its actions to protect its egg sacs are visible in the reconstructions.

Experiments to track other copepods as they reacted to changes in their environment were performed with T. Konotchick during the Celebes Cruise in October 2007 (Section 3.3.2) and made preliminary use of work developed during this thesis. The animals were located laterally using similar methods as Section 4.1.4. A 512×512 pixel region of the hologram around the copepod was extracted and the PFF response computed (Section 4.1.5), made more practical by the limited depth range and number of expected objects (one or two in all samples). The 3D path was estimated by combining the PFF responses and lateral positions into an energy model that promoted smooth motion, high PFF peak values, and consistent depth estimates [56]. The energy model was solved with simulated annealing (SA) and error bars estimated by running the SA multiple times. Figure 5-4 plots the estimated depth and the measured depth. For this example, the two depth estimates are close, and the regularization imposed through the energy actually helps smooth out the motion, though it does show sensitivity to the weights applied to each energy term. Accuracy for other copepods was lower because the PFF has wider and more peaks for real data, the sample tank partly obscured the view of the copepod when it was near the walls or bottom, and the motion model used was not able to adequately capture sudden jumps.

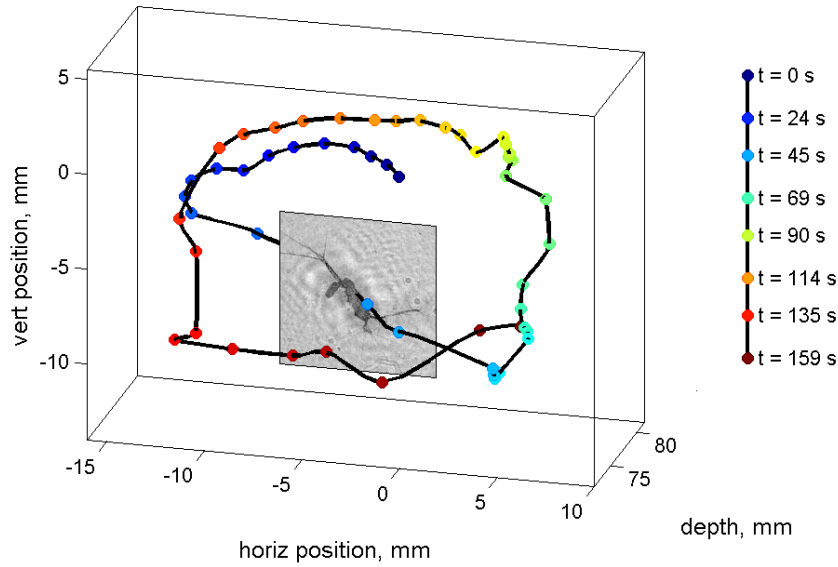


Figure 5-3: Example of tracking a female copepod along a 3D trail over time. The black line marks the path over time, with the reconstructed image showing that particular moment in time. The sample data was captured during the Celebes cruise (Section 3.3.2).

There are a number of improvements that can be made for future tracking experiments. A detector with a faster frame rate (or effective frame rate; see the next section) would reduce the uncertainty between time steps for how far the animal had moved. Switching to the MPFF (Section 4.1.5) or a brute-force reconstruction with the GPU would both generate better estimates of the depth. Finally, a motion model that allows for both slow drift and sudden jumps would significantly improve the solution for copepods. (Other animals with continuous motion have much simpler models.) Generating enough data to create the model and estimate its parameters would certainly be helped by the fast methods of Chapter 4.

5.2.4 Digital multiplexing holography

Section 3.1.2 showed that only a few digital bits, on the order of 4-5, are required to record a hologram with good reconstruction quality. Domínguez-Caballero and Loomis suggested that a substantial number of digital holograms could be recorded simultaneously in a single exposure, utilizing the excess bits in a sensor [99],[100]. Multiplexing holograms is a matter of pulsing the illumination multiple times during the camera's photosensitive period (Figure

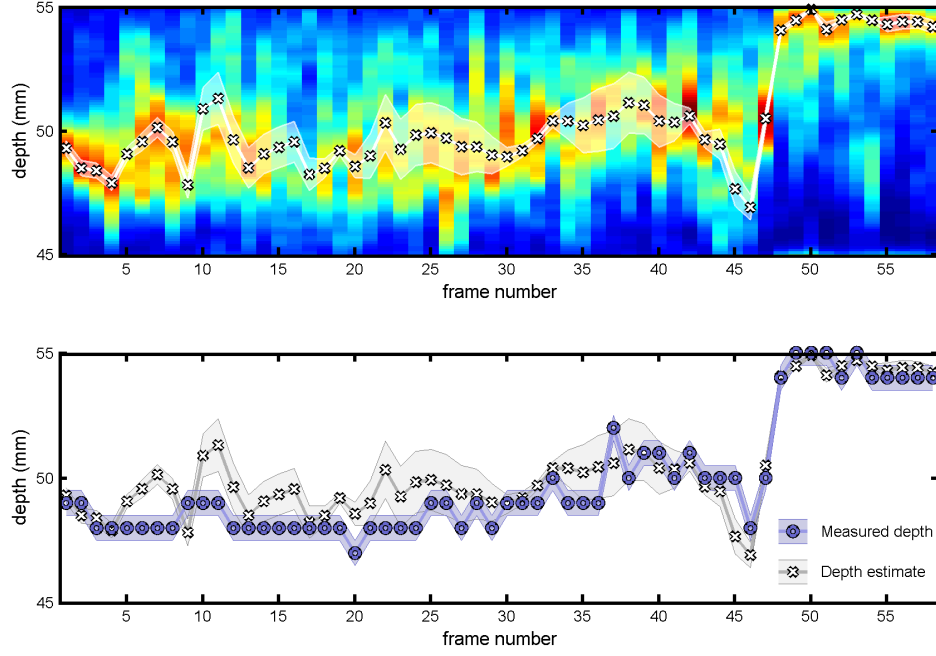


Figure 5-4: Using the PFF response to estimate a 3D motion path. Top: the PFF $F_c F_s$ is plotted vertically for each frame. Red indicates a peak while blue indicates a low value. The solution of the energy minimization is plotted as the “x” symbols with a thick white connector line, and includes both the PFF response, the lateral position, and a simple motion model. A $\pm 1\sigma$ error is plotted as the shaded area. Bottom: the estimated depth compared to the measured depth. The measured depth was judged to the nearest millimeter from reconstructed images. The copepod for this experiment was large, around $1200\ \mu\text{m}$, and spent some time near the edge and bottom of the tank where a significant fraction of the spatial frequencies were lost.

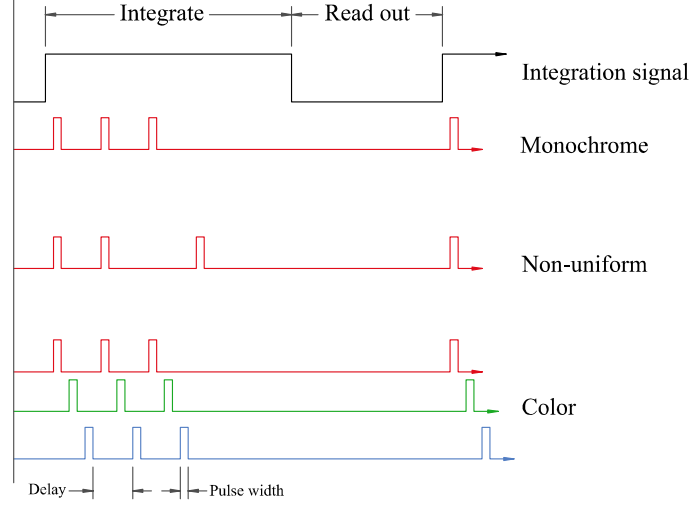


Figure 5-5: Three techniques for multiplexing digital holograms into a single exposure of a detector. Top: the camera is set to expose for an extended period, and the laser illumination is pulsed multiple times. Middle: the pulse rate is varied, providing an indication of the direction of motion if the objects have nearly constant velocity during the pulsing period. Bottom: multiple wavelengths are pulsed multiple times, disambiguating the direction of motion and providing higher temporal resolution. A camera with a color filter is required to take advantage of the third approach.

5-5), producing an incoherent sum of holograms,

$$I_M(x, y) = \sum_i I_i = \sum_i |r_i|^2 + \sum_i |o_i|^2 + \sum_i [r_i^* o + r_i o^*], \quad (5.1)$$

where I_i is the hologram corresponding to the i^{th} laser pulse and the r_i , o_i are the reference and object waves during the i^{th} period. The multiplexed hologram is reconstructed exactly like a normal hologram. The reconstruction is then the sum of the object images and may contain overlap if the objects have not shifted sufficiently between illumination pulses. Laboratory experiments were able to record 4-7 holograms within an exposure using a Kodak KAF-16801E 12-bit sensor, or up to 12 exposures with reduced contrast and resolution. An example of a copepod nauplii recorded with three pulses is shown in Figure 5-6.

Multiplexed digital holography has a number of potential applications. Domínguez-Caballero et al. suggested that the images could be used for tracking micro-organisms over space and time with high space-bandwidth product cameras. The high SBP allowed them to track a large volume of organisms simultaneously [99]. Salah et al. later showed an

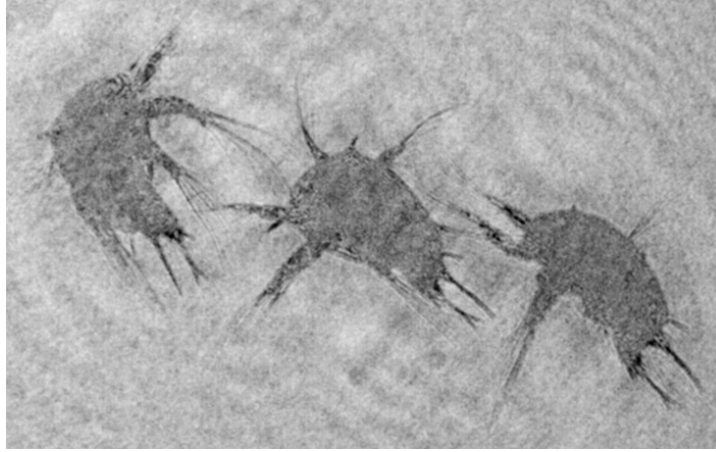


Figure 5-6: Multiplexed image of a copepod nauplii as it tumbled from left to right. The nauplii is estimated to be around $250\ \mu\text{m}$ and the image was recorded with a spherical reference.

example of tracking particles in a vortex street [310]. Other laboratory experiments have used multiplexing for acquiring bursts of particle images at equivalent frame rates of 4,200 fps, tracking fast-changing fluid flows with comparatively slow cameras.

Oceanic holography could benefit from multiplexing several volumes of water into a single hologram, effectively multiplying the total recorded volume per hologram. This assumes that the holographic device is moving through the water fast enough to image a new volume every few milliseconds. At a fast 12 knots, a 40 mm wide detector would require around 7 ms between pulses to completely separate each volume.

Another potentially interesting use might be for measuring marine microscale fluid flows. Malkiel et al. used digital holography in a laboratory setup to image tracer particles advected by a copepod's feeding flow [240], an experiment which may be more feasible to repeat *in situ* with fast multiplexing. Similarly, microscale turbulence and shearing [343],[305],[364] could be captured in a single snapshot if there are suitable tracers in the water. Alternately, streak velocimetry [94] and multiplexing could be combined to provide several velocity estimates within a single frame.

A caveat of multiplexed holography is that the noise builds quickly compared to the individual signals. This is a function of both the measurement noise [310] and the sheer number of particles [98],[97]. A low-noise sensor, such as a CCD, is thus critical for achieving

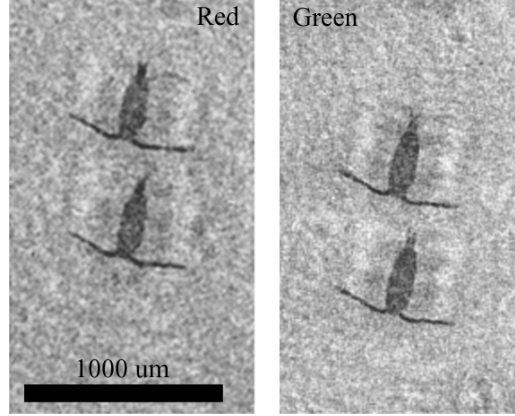


Figure 5-7: Example of combined time and spatial multiplexing. The red and green lasers were pulsed in sequence twice (see Figure 5-5). The reconstructions of the red-filtered pixels and green-filtered pixels are taken from the same area of the sensor, but are shown side-by-side for visual clarity. The copepod is moving unambiguously downward in these images. Note that without color multiplexing, the images would have overlapped due to the slow drift speed of the copepod.

higher numbers of multiplexed images.

Finally, it is worth noting that a number of methods for multiplexing are possible, three diagrammed in Figure 5-5. Time multiplexing is a good choice for a monochrome sensor. Color-based multiplexing is ideal for a sensor with a Bayer color filter, where two or more lasers record different images for each pulse. Figure 5-7 shows an example that utilizes both time and color multiplexing. Reference [99] discusses the use of polarization optics for recording multiple images. Both polarization and color multiplexing have the possibility of recording two different holograms simultaneously. For example, multiple views of the same object could be stored into one hologram. It should be noted that spatial multiplexing using color has appeared in the literature for both digital holography [207],[308] and quantitative phase retrieval [404],[405],[406],[229], so the concept is not necessarily novel but still useful. Precise measurements of phase objects may in fact benefit from these multiwavelength spatial multiplexing schemes, such as imaging *in situ* salt fingering.

5.2.5 Holography on gliders, drifters, and AUVs

The prototype unit has been deployed primarily on cabled platforms: CTD/rosettes, ROVs, and VPR drops. These platforms have a relatively slow speed through the water ($\lesssim 2$

m/s) and short deployments (1-3 hours). Three major challenges need to be examined in more detail before migrating the holography unit to more mobile platforms with longer deployments:

1. *Functional control.* The prototype is currently controlled via a programmable Micro PIC and stores holograms to an internal compact flash card. Increasing the memory card capacity and including an interface to trigger data capture would be the minimum required for simple AUV operations. A more flexible option, especially for advanced AUV missions, gliders, and drifters would be to replace the entire detector and control system with a low-power single-board computer and camera that interfaces directly with the computer.
2. *Power consumption.* The current prototype works from a single commercial battery pack, capturing hundreds of holograms on a single charge. The camera and laser are powered down when not in use, and the PIC uses a few milliwatts to monitor and control the device status. However, gliders and drifters are at sea for extended periods of time, and a certain amount of processing of the holograms would need to be done on board for the sake of conserving data storage and for efficiently transmitting only the data of interest. The methods of Chapter 4, specifically depth and position estimation (Sections 4.1.4 and 4.1.5), have the potential to vastly reduce the number of computations required and thus the energy expenditure.

Computational devices which use single instruction multiple data architectures, such as GPUs, are known to be particularly power-hungry in desktop applications. Fortunately, there has been considerable interest in reducing the size and power usage of GPUs for portable devices, and they are currently appearing in laptops and smart phones with greater frequency. It may well be worth investigating these portable GPUs for on-board processing: they are built purely for computation, so less overhead is wasted on background tasks and keeping the device active after the computations are finished¹.

¹Collange et al. note that desktop GPUs are around four times more efficient in terms of flops/watt than CPUs [70].

Optimizing algorithms will also require some thought towards power consumption. Collange et al. provide power consumption measurements of specific operations for a few select desktop GPUs. For example, reading from an L1 texture cache requires 3-4 times less power than global memory fetches, multiply-add (MAD) operations are 7-15 times more efficient than memory fetches, and reciprocals (RCP) require around three times the power of a MAD [70]. Thus, algorithms which compute kernels on-the-fly instead of reading them from memory might actually use less power. Similarly, boosting by interpolating from texture memory would be preferred over loading a classifier from global memory.

3. *Hydrodynamics.* Reducing drag is critical for gliders and drifters. A spherical reference allows a much smaller housing for the laser illumination. A camera with large space-bandwidth product and pixel size reduces the need for relay optics and similarly reduces the housing length. Fairings for the housings and struts are recommended.

Handling fast motion

The majority of the prototype deployments have been on slow moving platforms. As long as $v\tau \lesssim \Delta/M$, where v is the lateral velocity of an object, τ is the illumination pulse length, M is the magnification, and Δ is the effective pixel size (see Section 3.3.4), the motion blur will be nearly undetectable in the reconstructed image. For typical values of $M = 1.3$, $\tau = 5 \mu\text{s}$, and $\Delta = 15 \mu\text{m}$, the velocity should satisfy $v \lesssim 2.3 \text{ m/s}$. This estimate is rough, as some applications can withstand greater motion than others; for example, identification may not need (or be able to use) as high of a resolution as accurate particle sizing requires. A higher laser power and a more sensitive CCD can immediately increase the maximum velocity.

But what of the case where a device is pushed beyond its limit, either accidentally or intentionally? The effects of motion were previously discussed in Chapter 3 (see Section 3.1.1), with the assumption that most hologram motion was to be avoided. However, one of the main conclusions was that if there is significant lateral motion, the resulting hologram is nearly identical to the stationary hologram convolved with the motion PSF. This is exactly the same model as used for incoherent imaging and various well-known deconvolution

methods [31]—so that removing the motion blur is a matter of choosing a preferred method. In many traditional deconvolution methods, an initial guess for the motion PSF is necessary. Dixon et al. provide a way of estimating the direction and magnitude of short linear motion PSFs by looking at the directional blurring around a point particle’s diffraction image [94]. Oceanic holography is nearly guaranteed to have a number of particles in each frame to estimate the velocity.

One of the problems with linear motion blur is that the kernel includes zeros, destroying some image information and making the inversion unstable [31]. Raskar et al. have proposed modulating the illumination to create a kernel which has better inversion properties [2],[296]. Modulating the laser pulse is easy with a Micro PIC controller and a laser diode, both already used on the prototype. Initial experiments show that standard deconvolution with a modulated blur PSF improves visual image quality over deconvolution with a uniform blur PSF. Selecting an optimal modulation depends on the expected velocity, total exposure time, and the maximum modulation rate of the laser diode driver.

5.2.6 Riesz-based focus metric

Kovesi developed a measure of the phase alignment of Fourier components as a way of detecting edges in images, termed the “phase congruency”. The phase congruency measure has a number of benefits, including single peaks for bar-like shapes (such as antennae) and a lower dependence on the total contrast [201],[200],[199]. Unfortunately, computing the phase congruency directly is relatively slow. Venkatesh and Owens note that the peaks of the phase congruency are at the same points as the peaks of the local energy [199], which can be measured as the amplitude of the Riesz transform. In a sense, the Riesz amplitude is a proxy for the phase congruency. Figure 5-8 shows an example of using the Riesz amplitude as a focus metric. Given that the Riesz has intrinsic connections to local frequency and can be computed for complex-valued fields, it is a particularly interesting candidate for future research.

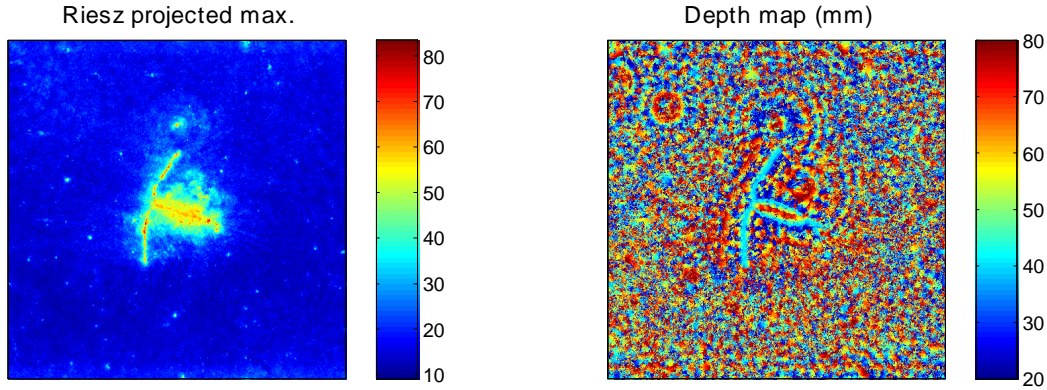


Figure 5-8: The amplitude of the Riesz transform used as a focus measure, projected in the same manner as in Section 4.1. This figure uses the same example hologram as Figure 4-8. The copepod is in focus at $z = 40.6$ mm, which corresponds well to the teal outline of the copepod visible in the depthmap.

5.3 Future outlook

Digital imaging devices have advanced significantly over the past decade. High space-bandwidth product cameras, both CMOS and CCD, are becoming commonplace on the market, and demand for faster frame rates makes generating enormous amounts of visual data simple. Making sense of the information and managing massive data sets is becoming the next need for science. Reliable, accurate, and fast automated methods for extraction and analysis are a necessity.

The past decade has also seen an explosive increase in computational power. Desktop CPUs are necessarily going to improve, but the biggest changes have been in mobile devices and GPUs. Single computers built around general purpose GPU processing now boast computational speeds that rival the top supercomputers of the late 1990's and early 2000's. Mobile phones have faster processing with lower power consumption than laptops just a few years previous. The potential for extremely fast laboratory processing, shipboard “supercomputers”, and efficient deployable platforms is incredible. The computational ability is approaching levels where it is also not a limiting factor.

Fortunately for biology and ecology, the ability to compute is leading computer scientists to create incredibly potent new methods. It is a matter of deciding which of these can be adapted for practical use in a challenging and noise-filled ocean environment. As demon-

strated in this thesis, modifications and novel approaches are always possible, inspired by previous work.

Digital holography is at an interesting junction between physics and computer science. The foundational ideas have been well studied and reported in the literature. The basic tools have been created and analyzed. New approaches will center around working with the images generated by digital holography. Novel advances are decidedly possible by combining previous computer vision work with complex-valued three-dimensional optical fields, a dataset that is itself decidedly different from the images most computer scientists are used to considering.

At the end of a long day, the message is that biological imaging is at a fantastic cusp in its development as an applicable tool for real-world science.



Figure 5-9: Example of iterated local clustering based on color similarity. A 2D map of high-dimensional color data is first divided into vertical strips. Within each strip, horizontal scan lines are clustered and the statistics computed. The two clusters with the shortest distance between their mean values are combined, then the statistics re-computed for the new cluster. The process is iterated until the distance between all the centers are greater than a specified threshold. For visualization, the size of the clusters has been jittered slightly.

Bibliography

- [1] Abramowitz, M. and I. Stegun. *Handbook of Mathematical Functions*. Dover Publications Inc., 1965.
- [2] Agrawal, Amit and Ramesh Raskar. “Resolving Objects at Higher Resolution from a Single Motion-blurred Image.” *Proc. IEEE Computer Vision and Pattern Recognition*. 2007.
- [3] Agrawal, Y. C. and H. C. Pottsmith. “Instruments for particle size and settling velocity observations in sediment transport,” *Marine Geology*, 168:89–114 (2000).
- [4] Aherne, F.J., et al. “The Bhattacharyya metric as an absolute similarity measure for frequency coded data,” *Kybernetika*, 32:363–368 (1997).
- [5] Ahn, Jong Ho and Stanley B. Grant. “Size Distribution, Sources, and Seasonality of Suspended Particles in Southern California Marine Bathing Waters,” *Environmental Science and Technology*, 41:695–702 (2007).
- [6] Aiken, J., et al. “The measurement of optical attenuation in sea water.” *Proceedings of IEEE OCEANS* 2. 689–693. 1994.
- [7] Akaike, H. “A new look at the statistical model identification,” *IEEE Trans. on Automatic Control*, 19:716–723 (1974).
- [8] Allen, J. B. and L. R. Rabiner. “A unified approach to short-time Fourier analysis and synthesis,” *Proceedings of the IEEE*, 65:1558–1564 (1977).

- [9] Alvain, S., et al. “Remote sensing of phytoplankton groups in case 1 waters from global SeaWiFS imagery,” *Deep-Sea Research Part I: Oceanographic Research Papers*, 52:1989–2004 (2005).
- [10] American Society for Testing and Materials, “ASTM G173-03 Standard Tables for Reference Solar Spectral Irradiances: Direct Normal and Hemispherical on 37 degree Tilted Surface.” ASTM International, West Conshohocken, PA, 2008.
- [11] Antkowiak, Maciej, et al. “Extended focused imaging of a microparticle field with digital holographic microscopy,” *Optics Letters*, 33:1626–1628 (2008).
- [12] Arfken, George B. and Hans J. Weber. *Mathematical Methods for Physicists*. Harcourt Academic Press, 2001.
- [13] Asundi, Anand and Vijay Raj Singh. “Sectioning of amplitude images in digital holography,” *Measurement Science and Technology*, 17:75–78 (2006).
- [14] Atherton, T. J. and D. J. Kerbyson. “Size invariant circle detection,” *Image and Vision Computing*, 17:795–803 (1999).
- [15] Babin, Marcel, et al. “Light scattering properties of marine particles in coastal and open ocean waters as related to particle mass concentration,” *Limnol. Oceanogr.*, 48:843–859 (2003).
- [16] Balboa, Rosario M. and Norberto M. Grzywacz. “Power spectra and distribution of contrasts of natural images from different habitats,” *Vision Research*, 43:2527–2537 (2003).
- [17] Barsi, Christopher, et al. “Imaging through nonlinear media using digital holography,” *Nature Photonics*, 3:211–215 (2009).
- [18] Barwick, D. Shane. “Very Fast Best-Fit Circular and Elliptical Boundaries by Chord Data,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31:1147–1152 (2009).
- [19] Bastiaans, M. J. “Wigner distribution function and its application to first-order optics,” *Journal of the Optics Society of America*, 69:1710–1716 (1979).

- [20] Bastiaans, M. J. “Application of the Wigner distribution function in optics.” *The Wigner Distribution - Theory and Applications in Signal Processing* edited by W. Mecklenbrauker and F. Hlawatsch, 375–426, Elsevier Science, Amsterdam, 1997.
- [21] Beaugrand, Grégory, et al. “Plankton effect on cod recruitment in the North Sea,” *Nature*, 426:661–664 (2003).
- [22] Beaugrand, Grégory and Philip C. Reid. “Long-term changes in pytoplankton, zoo-plankton, and salmon related to climate,” *Global Change Biology*, 9:801–817 (2003).
- [23] Beckman Coulter, Inc., “Z Series Coulter Counters.” Online at <http://www.coulterflow.com/bciflow/instruments12.php>, December 8 2010.
- [24] Belhumeur, Peter N., et al. “Eigenfaces vs Fisherfaces: Recognition Using Class Specific Linear Projection,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19:711–720 (1997).
- [25] Belkasim, S.O., et al. “Pattern recognition with moment invariants: A comparative study and new results,” *Pattern Recognition*, 24:1117–1138 (1991).
- [26] Benfell, Carol, “Fishermen Oppose New Off-limits Zones, Overflow Crowd in Sebastopol Protests During Fish and Game’s Initial Marine Protection Meeting.” Press Democrat, Santa Rosa CA, July 10 2001.
- [27] Benfell, Carol, “State Plans to Close Rockfish Season Friday, October 5, 2001: Fish and Game officials say two species are depleted, forcing second closure this year.” Press Democrat, Santa Rosa, October 4 2001.
- [28] Benfield, M. C., et al. “ZOOVIS: a high resolution digital camera system for quantifying zooplankton abundance and environmental data.” *American Society of Limnology and Oceanography Aquatics Sciences Meeting, Albuquerque, NM*. 2001.
- [29] Benfield, Mark C., et al. “RAPID, Research on Automated Plankton Identification,” *Oceanography*, 20:172–187 (2007).

- [30] Benfield, Mark C., et al. *Handbook of Scaling Methods in Aquatic Ecology*, chapter Measurement of Zooplankton Distributions with a High-Resolution Digital Camera System, 17–30. CRC Press, 2004.
- [31] Bertero, Mario and Patrizia Boccacci. *Introduction to Inverse Problems in Imaging*. Institute of Physics Publishing, Bristol, 2002.
- [32] Bertsimas, Dimitris and John N. Tsitsiklis. *Introduction to Linear Optimization*. Athena Scientific, 1997.
- [33] Bhattacharyya, A. “On a Measure of Divergence Between Two Statistical Populations Defined by their Probability Distributions,” *Bull. Calcutta Math. Soc.*, 35:99–110 (1943).
- [34] Bilsky, Eric A. “Conserving Marine Habitats,” *Sustainable Development Law & Policy*, 84:67–70 (2006).
- [35] Borda, Monica, et al. “New instantaneous frequency estimation method based on image processing techniques,” *Journal of Electronic Imaging*, 14:023013 (2005).
- [36] Born, Max and Emil Wolf. *Principles of Optics*. Cambridge University Press, 1999.
- [37] Bouman, Charles A., “Cluster: An Unsupervised Algorithm for Modeling Gaussian Mixtures.” Available at <https://engineering.purdue.edu/bouman/software/cluster/>, October 10 2008.
- [38] Boyce, Daniel G., et al. “Global phytoplankton decline over the past century,” *Nature*, 466:591–596 (2010).
- [39] Boykov, Y. and G. Funka-Lea. “Graph cuts and efficient N-D image segmentation,” *International Journal of Computer Vision*, 70:109–131 (2006).
- [40] Brady, David J., et al. “Compressive Holography,” *Optics Express*, 17:13040–13049 (2009).

- [41] Brothers, N. P., et al. “The Incidental Catch of Seabirds by Longline Fisheries: Worldwide Review and Technical Guidelines for Mitigation,” *FAO Fisheries Circular*, 937 (1999).
- [42] Broughton, Elisabeth A. and R. Gregory Lough. “A direct comparison of MOCNESS and Video Plankton Recorder zooplankton abundance estimates: Possible applications for augmenting net sampling with video systems,” *Deep-Sea Research Part II: Topical Studies in Oceanography*, 53:2789–2807 (2006).
- [43] Brown, Matthew and David G. Lowe. “Automatic Panoramic Image Stitching using Invariant Features,” *International Journal of Computer Vision*, 74:59–73 (2007).
- [44] Buiteveld, H., et al. “The optical properties of pure water.” *SPIE Proceedings on Ocean Optics* 2258. 174–183. 1994.
- [45] Burghouts, G. and J.M. Geusebroek. “Color Textons for Texture Recognition,” *British Machine Vision Conference*, 3:1099–1108 (2006).
- [46] Burnham, Kenneth P. and David R. Anderson. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. Springer, 2002.
- [47] Buskey, E. J., et al. “Escape behavior of planktonic copepods to hydrodynamic disturbances: high speed video analysis,” *Marine Ecology Progress Series*, 235:135–146 (2002).
- [48] Buskey, Edward J. and Daniel K. Hartline. “High-Speed Video Analysis of the Escape Responses of the Copepod *Acartia tonsa* to Shadows,” *Biological Bulletin*, 204:28–37 (2003).
- [49] California Department of Fish and Game, “DFG Implements Emergency Sport Fishing Closure for Offshore Rockfish.” Press Release, November 7 2001.
- [50] California Department of Fish and Game, “DFG Recommends Emergency Closure of Offshore Rockfish and Longcod Sport Fisheries.” Press Release, October 2 2001.
- [51] California Department of Fish and Game, Marine Region, “Status of the Fisheries Report, An Update Through 2008,” August 2010.

- [52] Camilli, R., et al. “Rapid Swath Mapping of Reef Ecology and Associated Water Column Chemistry in the Gulf of Chiriqui, Panama.” *Proc. IEEE Oceans*. 2007.
- [53] Camilli, Richard, et al. “Tracking Hydrocarbon Plume Transform and Biodegradation at Deepwater Horizon,” *Science*, 330:201–204 (2010).
- [54] Canny, J. “A computational approach to edge detection,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8:679–698 (1986).
- [55] Caron, Nicolas and Yunlong Sheng. “Polynomial phase masks for extending the depth of field of a microscope,” *Applied Optics*, 47:E39–E43 (2008).
- [56] Chalmond, Bernard. *Modeling and Inverse Problems in Image Analysis*. Springer, 2003.
- [57] Chan, Tony F. and Luminita A. Vese. “Active Contours Without Edges,” *IEEE Transactions on Image Processing*, 10:266–277 (2001).
- [58] Charette, Matthew and Walter H. F. Smith. “The Volume of Earth’s Ocean,” *Oceanography*, 23:112–114 (2010).
- [59] Checkley, D. M., Jr., et al. “Assessing plankton and other particles in situ with the SOLOPC,” *Limnology and Oceanography*, 53:2123–2136 (2008).
- [60] Checkley, David M, Jr. “Fisheries Management and Ocean Observations.” *OceanObs*. 2009.
- [61] Cheng, Yizong. “Mean Shift, Mode Seeking, and Clustering,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17:790–799 (1995).
- [62] Cheng, Yu Chin and Samuel C. Lee. “A new method for quadratic curve detection using k-RANSAC with acceleration techniques,” *Pattern Recognition*, 28:663–682 (1995).
- [63] Choksuriwong, A., et al. “Object Recognition Using Local Characterisation and Zernike Moments,” *Advanced Concepts for Intelligent Vision Systems, Lecture Notes in Computer Science*, 3708:108–115 (2005).

- [64] Chong, C. W., et al. “Translation and Scale Invariants of Legendre Moments,” *Pattern Recognition*, 37:119–129 (2004).
- [65] Christiano, P., et al., “Electrical Flows, Laplacian Systems, and Faster Approximation of Maximum Flow in Undirected Graphs.” preprint available at arxiv.org/pdf/1010.2921, February 23 2011.
- [66] Chuang, Y. Y., et al. “A Bayesian Approach to Digital Matting,” *Proceedings of the IEEE Computer Vision and Pattern Recognition*, 2:264–271 (2001).
- [67] Chung, Kuo-Liang, et al. “Fast randomized algorithm for center-detection,” *Pattern Recognition*, 43:2659–2665 (2010).
- [68] Ciotti, Aurea M. and Annick Bricaud. “Retrievals of a size parameter for phytoplankton and spectral light absorption by colored detrital matter from water-leaving radiances at SeaWiFS channels in a continental shelf region off Brazil,” *Limnology and Oceanography: Methods*, 4:237–253 (2006).
- [69] Cohen, Jonathan H. and Jr. Richard B. Forward. “Spectral Sensitivity of Vertically Migrating Marine Copepods,” *Biological Bulletin*, 203:307–314 (2002).
- [70] Collange, Sylvain, et al. “Power Consumption of GPUs from a Software Perspective.” *Proceedings of the 9th International Conference on Computer Science*. 2009.
- [71] Comaniciu, D., et al. “The variable bandwidth mean shift and data-driven scale selection,” *Proceedings of the International Conference on Computer Vision*, 1:438–445 (2001).
- [72] Comaniciu, Dorin and Peter Meer. “Mean Shift: A Robust Approach Toward Feature Space Analysis,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24:603–619 (2002).
- [73] Committee on Understanding Oil Spill Dispersants: Efficacy and Effects. *Oil spill dispersants, efficacy and effects*. The National Academic Press, Washington DC, 2005.

- [74] Conover, Robert J., et al. “Climate, copepods, and cod: some thoughts on the long-range prospects for a sustainable northern cod fishery,” *Climate Research*, 5:69–82 (1995).
- [75] Coppola, G., et al. “A digital holographic microscope for complete characterization of microelectricalmechanical systems,” *Measurement Science and Technology*, 15:529–539 (2004).
- [76] Cormen, Thomas H., et al. *Introduction to Algorithms*. MIT Press, 2009.
- [77] Cornelis, N. and L. Van Gool. “Fast Scale Invariant Feature Detection and Matching on Programmable Graphics Hardware.” *Proc. IEEE Computer Vision and Pattern Recognition Workshop*. 2008.
- [78] Costa, Amy DeLorenzo, et al. “Variability in the nutritional value of the major copepods in Cape Code Bay (Massachusetts, USA) with implications for right whales,” *Marine Ecology*, 27:109–123 (2006).
- [79] Cowen, R. K. and C. M. Guigand. “*In situ* ichthyoplankton imaging system (ISIIS): system design and preliminary results,” *Limnology and Oceanography: Methods*, 6:126–132 (2008).
- [80] Cox, Michael A. A. and Trevor F. Cox. “Multidimensional Scaling.” *Handbook of Data Visualization* Springer Berlin, 2008.
- [81] Culverhouse, Phil, et al. “Automatic image analysis of plankton: future perspectives,” *Marine Ecology Progress Series*, 312:297–309 (2006).
- [82] Dahms, Hans-Uwe and Jiang-Shiou Hwang. “Perspectives of underwater optics in biological oceanography and plankton ecology studies,” *Journal of Marine Science and Technology*, 18:112–121 (2010).
- [83] Darakis, E., et al. “Microparticle characterization using digital holography,” *Chemical Engineering Science*, 65:1037–1044 (2010).
- [84] Davis, C. S., et al. “Microaggregations of oceanic plankton observed by towed video microscopy,” *Science*, 257:230–232 (1992).

- [85] Davis, Cabell, et al. “The Video Plankton Recorder (VPR): design and initial results,” *Arch. Hydrobiol. Beiheft Ergeb. Limnol.*, 36:67–81 (1992).
- [86] Davis, Cabell S., et al. “Rapid visualization of plankton abundance and taxonomic composition using the Video Plankton Recorder,” *Deep-Sea Research II*, 43:1947–1970 (1996).
- [87] Davis, Cabell S. and Dennis J. McGillicuddy. “Transatlantic Abundance of the N_2 -Fixing Colonial Cyanobacterium *Trichodesmium*,” *Science*, 312:1517–1520 (2006).
- [88] Davis, Cabell S., et al. “A three-axis fast-tow digital Video Plankton Recorder for rapid surveys of plankton taxa and hydrography,” *Limnology and Oceanography: Methods*, 3:59–74 (2005).
- [89] de Jong, Jeremy and Hui Meng. “Digital holographic particle validation via complex wave,” *Applied Optics*, 46:7652–7661 (2007).
- [90] De Nicola, S., et al. “Angular spectrum method with correction of anamorphism for numerical reconstruction of digital holograms on tilted planes,” *Optics Express*, 13:9935–9940 (2005).
- [91] Delen, N. and B. Hooker. “Free-space beam propagation between arbitrary oriented planes based on full diffraction theory: a fast Fourier transform approach,” *Journal of the Optics Society of America A*, 15:857–867 (1998).
- [92] Denis, Loïc, et al. “Direct extraction of the mean particle size from a digital hologram,” *Applied Optics*, 45:944–952 (2006).
- [93] Denman, Kenneth, et al. “Observing planktonic ecosystems: needs, capabilities, and a strategy for the next decade.” *OceanObs09*. 2009.
- [94] Dixon, Lisa, et al. “Holographic particle-streak velocimetry,” *Optics Express*, 19:4393–4398 (2011).
- [95] Domínguez-Caballero, José A. *Digital Holographic Imaging of Aquatic Species*. MS thesis, Massachusetts Institute of Technology, 2006.

- [96] Domínguez-Caballero, José A. *Optimization of the Holographic Process for Imaging and Lithography*. PhD dissertation, Massachusetts Institute of Technology, 2009.
- [97] Domínguez-Caballero, José A. and G. Barbastathis. “Stability of Inversion in Digital Holographic Particle Imaging: Theory and Experimental Validation.” *Frontiers in Optics, Rochester, NY, USA*. 2008.
- [98] Domínguez-Caballero, José A. and G. Barbastathis. “Stability of the Digital Holographic Inverse Problem as a Function of Particle Density.” *OSA Conference on Digital Holography and Three-Dimensional Imaging, St. Petersburg, FL, USA*. 2008.
- [99] Domínguez-Caballero, José A., et al. “Techniques Based on Digital Multiplexing Holography for Three-Dimensional Object Tracking.” *Conference on Lasers and Electro-Optics*. 2007.
- [100] Domínguez-Caballero, José A., et al. “Advances in Plankton Imaging using Digital Holography.” *OSA Conference on Digital Holography and Three-Dimensional Imaging, Vancouver, BC, Canada*. 2007.
- [101] Donoser, M. and H. Bischof. “3D Segmentation by Maximally Stable Volumes,” *Proceedings of the International Conference on Pattern Recognition*, 63–66 (2006).
- [102] Dowski, E. R. and W. T. Cathey. “Extended depth of field through wave-front encoding,” *Applied Optics*, 34:1859–1866 (1995).
- [103] Dubois, F., et al. “Focus plane detection criteria in digital holography microscopy by amplitude analysis,” *Optics Express*, 14:5895–5908 (2006).
- [104] Dubois, Frank, et al. “Improved three-dimensional imaging with a digital holographic microscope with a source of partial spatial coherence,” *Applied Optics*, 38:7085–7094 (1999).
- [105] Dubois, Frank, et al. “Partial spatial coherence effects in digital holographic microscopy with a laser source,” *Applied Optics*, 43:1131–1139 (2004).
- [106] Duda, Richard O., et al. *Pattern Classification*. John Wiley and Sons, 2001.

- [107] Eastman Kodak Company, “KAF-16801E Performance Specification,” August 22 2000.
- [108] Eastman Kodak Company, “Kodak KAF-16802 Image Sensor, device performance specification revision 2.0 MTD/PS-0528,” February 2007.
- [109] Eastman Kodak Company, “Kodak KAF-16803 Image Sensor, Device Performance Specification,” August 10 2010.
- [110] Eastman Kodak Company, “Kodak KAF-39000 Image Sensor, Device Performance Specification,” August 31 2010.
- [111] Edwards, Martin and Anthony J. Richardson. “Impact of climate change on marine pelagic phenology and trophic mismatch,” *Nature*, *420*:881–884 (2004).
- [112] Fabry, Victoria J. “Marine Calcifiers in a High- CO_2 Ocean,” *Science*, *23*:1020–1022 (2008).
- [113] Fang, Xiao and Peter R. Hobson. “Effect of spherical aberration on real-image fidelity from replayed in-line holograms of underwater objects,” *Applied Optics*, *37*:3206–3214 (1998).
- [114] Felsberg, Michael and Gerald Sommer. “The Monogenic Signal,” *IEEE Transactions on Signal Processing*, *49*:3136–3144 (2001).
- [115] Ferraro, P., et al. “Controlling image size as a function of distance and wavelength in Fresnel-transform reconstruction of digital holograms,” *Optics Letters*, *15*:854–856 (2004).
- [116] Fienup, J. R. “Invariant error metrics for image reconstruction,” *Applied Optics*, *32*:8352–57 (1997).
- [117] Fischler, M. A. and R. C. Bolles. “Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography,” *Comm. Assoc. Comp. Mach.*, *24*:381–395 (1981).

- [118] Flusser, J. and T. Suk. “Rotation Moment Invariants for Recognition of Symmetric Objects,” *IEEE Transactions on Image Processing*, 15:3784–3790 (2006).
- [119] Food and Agriculture Organization, United Nations, “The State of World Fisheries and Aquaculture 2008,” 2009.
- [120] Food and Agriculture Organization, United Nations, “The State of World Fisheries and Aquaculture 2010,” 2011.
- [121] Foote, K. G. and T. K. Stanton. *ICES Zooplankton Methodology Manual*, chapter Acoustical methods, 223–258. Academic Press, London, 2000.
- [122] Forssén, Per-Erik. “Maximally Stable Colour Regions for Recognition and Matching.” *Computer Vision and Pattern Recognition*. 2007.
- [123] Forward, R. B. “Diel vertical migration: Zooplankton photobiology and behaviour,” *Oceanography and Marine Biology - An Annual Review*, 26:361–393 (1988).
- [124] Francis, Malcolm P., et al. “Pelagic shark bycatch in the New Zealand tuna longline fishery,” *Marine and Freshwater Research*, 52:165–178 (2001).
- [125] Freedman, Daniel and Tao Zhang. “Interactive Graph Cut Based Segmentation With Shape Priors.” *Computer Vision and Pattern Recognition*. 755–762. 2005.
- [126] Freeman, William T. and Edward H. Adelson. “The Design and Use of Steerable Filters,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13:891–906 (1991).
- [127] Freund, Yoav and Robert E. Schapire. “A Short Introduction to Boosting,” *Journal of the Japanese Society for Artificial Intelligence*, 14:771–780 (1999).
- [128] Frey, Brendan J. and Delbert Dueck. “Clustering by Passing Messages Between Data Points,” *Science*, 315:972–976 (2007).
- [129] Friedman, Jerome, et al. “Additive Logistic Regression: a Statistical View of Boosting,” *The Annals of Statistics*, 28:337–374 (2000).

- [130] Frosio, I. and N. A. Borghese. “Real-time accurate circle fitting with occlusions,” *Pattern Recognition*, 41:1041–1055 (2008).
- [131] Fugal, J. P., et al. “Practical methods for automated reconstruction of particles in digital in-line holograms,” *Measurement Science and Technology*, 20:075501 (2009).
- [132] Fuller, Susanna D., et al., “How We Fish Matters: Addressing the Ecological Impacts of Canadian Fishing Gear.” Ecology Action Centre, Living Oceans Society, and Marine Conservation Biology Institute, 2009.
- [133] Gaedke, Ursula. “The size distribution of plankton biomass in a large lake and its seasonal variability,” *Limnology and Oceanography*, 37:1202–1220 (1992).
- [134] Gallager, Scott M., et al. “Contribution of fine-scale vertical structure and swimming behavior to formation of plankton layers on Georges Bank,” *Marine Ecology Progress Series*, 267:27–43 (2004).
- [135] Garcia, V., et al. “Fast k nearest neighbor search using GPU.” *Proc. IEEE Computer Vision and Pattern Recognition, Anchorage, AK*. 1–6. 2008.
- [136] Garcia-Sucerquia, J., et al. “Immersion digital in-line holographic microscopy,” *Optics Letters*, 31:1211–1213 (2006).
- [137] Garcia-Sucerquia, Jorge, et al. “Digital in-line holographic microscopy,” *Applied Optics*, 45:836–850 (2006).
- [138] Gävert, Hugo, et al., “FastICA 2.5 for Matlab.” Laboratory of Computer and Information Science, Adaptive Informatics Research Centre, Helsinki University of Technology. Available at <http://www.cis.hut.fi/projects/ica/fastica/code/dlcode.shtml>.
- [139] Georgescu, B., et al. “Mean shift based clustering in high dimensions: a texture classification example,” *Proceedings of the International Conference on Computer Vision*, 1:456–463 (2003).
- [140] Gersho, Allen and Robert M. Gray. *Vector quantization and signal compression*. Kluwer Academic Press, 1992.

- [141] Ghorayeb, Hicham, et al. *Lecture Notes in Computer Science*, 3852, chapter Boosted Algorithms for Visual Object Detection on Graphics Processing Units, 254–263. Springer-Verlag, Berlin, 2006.
- [142] Gills, Justin, “Extreme Heat Bleaches Coral, and Threat Is Seen.” *New York Times*, September 20 2010.
- [143] Gitay, Habiba, et al., editors. *Climate Change and Biodiversity, IPCC Technical Paper V*. Intergovernmental Panel on Climate Change, 2002.
- [144] Goldberg, Andrew V. and Robert E. Tarjan. “A new approach to the maximum-flow problem,” *Journal of the ACM*, 35:921–940 (1988).
- [145] Goodman, Joseph. *Introduction to Fourier Optics*. Roberts and Company Publishers, Greenwood Village, 2005.
- [146] Gorsky, G., et al. “Use of the underwater video profilers for the study of aggregate dynamics in the North Mediterranean,” *Estuarine, Coastal and Shelf Science*, 50:121–128 (2000).
- [147] Gorsky, Gaby, et al. “Digital zooplankton image analysis using the ZooScan integrated system,” *Journal of Plankton Research*, 32:285–303 (2010).
- [148] Goswami, Jaideva C. and Albert E. Hoefel. “Algorithms for estimating instantaneous frequency,” *Signal Processing*, 84:1423–1427 (2004).
- [149] Goudail, F. and P. Réfrégier. “Optimal target tracking on image sequences with a deterministic background,” *Journal of the Optics Society of America A*, 14:3197–3207 (1997).
- [150] Guo, Cheng-Shan, et al. “Effect of the fill factor of CCD pixels on digital holograms: comment of the papers ‘Frequency analysis of digital holography’ and ‘Frequency analysis of digital holography with reconstruction by convolution’,” *Optical Engineering*, 42:2768–2771 (2003).
- [151] Hahn, Joonku, et al. “Compressive Holographic Microscopy.” *Optical Society of America Biomedical Optics, Miami, Florida*. 2010.

- [152] Hall, Mark, et al. "The WEKA Data Mining Software: An Update," *SIGKDD Explorations*, 11 (2009).
- [153] Harris, Chris and Mike Stephens. "A Combined Corner and Edge Detector." *Proceedings of the 4th Alvey Vision Conference*. 147–152. 1988.
- [154] Hays, Graeme C., et al. "Climate change and marine plankton," *Trends in Ecology and Evolution*, 20:337–344 (2005).
- [155] Hecht, Eugene. *Optics*. Addison Wesley Publishing, 1997.
- [156] Hennelly, B. and J. Sheridan. "Generalizing, optimizing, and inventing numerical algorithms for the fractional Fourier, Fresnel, and linear canonical transforms," *Journal of the Optics Society of America A*, 22:917–927 (2005).
- [157] Herbland, A., et al. "Size structure of phytoplankton biomass in the equatorial Atlantic Ocean," *Deep-Sea Research*, 32:819–836 (1985).
- [158] Herman, A. W. "Design and calibration of a new Optical Plankton Counter capable of sizing small zooplankton," *Deep-Sea Research*, 39:395–415 (1992).
- [159] Herman, A. W., et al. "The next generation of the Optical Plankton Counter: the Laser-OPC," *Journal of Plankton Research*, 26:1135–1145 (2004).
- [160] Hinsch, K. D. "Holographic particle image velocimetry," *Measurement Science and Technology*, 13:R61–R72 (2002).
- [161] Hobson, Peter R. and John Watson. "The principles and practice of holographic recording of plankton," *Journal of Optics A: Pure and Applied Optics*, 4:S34–S49 (2002).
- [162] Hooff, Rian C. and William T. Peterson. "Copepod biodiversity as an indicator of changes in ocean and climate conditions of the northern California current ecosystem," *Limnology and Oceanography*, 51:2607–2620 (2006).
- [163] Hotelling, Harold. "The Generalization of Student's Ratio," *The Annals of Mathematical Statistics*, 2:360–378 (1931).

- [164] Hu, Ming-Kuei. “Visual pattern recognition by moment invariants,” *IRE Trans. Information Theory*, 8:179–187 (1962).
- [165] Hu, Qiao. *Application of Statistical Learning Theory to Plankton Image Analysis*. PhD dissertation, Massachusetts Institute of Technology and Woods Hole Oceanographic Institute, 2006.
- [166] Hu, Qiao and Cabell Davis. “Accurate automatic quantification of taxa-specific plankton abundance using dual classification with correction,” *Marine Ecology Progress Series*, 306:51–61 (2006).
- [167] Huss, G., et al. “Spatial filtering efficiency of single-mode optical fibers for stellar interferometry applications: phenomenological and numerical study,” *Optics Communications*, 244:209–217 (2005).
- [168] Hwang, Jiang-Shiou and Rudi Strickler. “Can Copepods Differentiate Prey from Predator Hydromechanically?,” *Zoological Studies*, 40:1–6 (2001).
- [169] Hyvärinen, A. “Fast and Robust Fixed-Point Algorithms for Independent Component Analysis,” *IEEE Trans. on Neural Networks*, 10:626–634 (1999).
- [170] Hyvärinen, Aapo, et al. *Independent Component Analysis*. Wiley-Interscience, 2001.
- [171] Jackson, J. B. C., et al. “Historical Overfishing and the Recent Collapse of Coastal Ecosystems,” *Science*, 293:629–638 (2001).
- [172] Jaffe, J. S. “Sensing Plankton: Acoustics and Optical Imaging.” *European Workshop on Harmful Algal Blooms*. 2005.
- [173] Jaffe, Jules S., et al. “Underwater Optical Imaging: Status and Prospects,” *Oceanography*, 14:64–75 (2001).
- [174] Jähne, Bernd. *Digital Image Processing*. Springer, 2005.
- [175] Janesick, James R. *Scientific Charge-Coupled Devices*. SPIE Press, 2001.

- [176] Javidi, Bahram, et al. "Three-dimensional imaging and recognition of microorganism using single-exposure on-line (SEOL) digital holography," *Optics Express*, 13:4492–4506 (2005).
- [177] Javidi, Bahram and Enrique Tajahuerce. "Three-dimensional object recognition by use of digital holography," *Optics Letters*, 25:610–612 (2000).
- [178] Javidi, Bahram, et al. "Real-time automated 3D sensing, detection, and recognition of dynamic biological micro-organic events," *Optics Express*, 14:3806–3829 (2006).
- [179] Jericho, S, K., et al. "Submersible digital in-line holographic microscope," *Review of Scientific Instruments*, 77:043706 (2006).
- [180] Johnson, Don H. and Sinan Sinanovic. "Symmetrizing the Kullback-Leibler Distance," *IEEE Transactions on Information Theory* (2001).
- [181] Johnson, Douglas H., et al. "Incidental catch of marine birds in the North Pacific high seas driftnet fisheries in 1990," *International North Pacific Fisheries Commission Bulletin*, 53:473–483 (1993).
- [182] Joint Analysis Group, "Review of Preliminary Data to Examine Subsurface Oil in the Vicinity of MC252-1, May 19 to June 19, 2010." Available from NOAA at <http://beta.w1.noaa.gov/sciencemissions/>, 2010.
- [183] Kan, Chao and Mandyam D. Srinath. "Invariant character recognition with Zernike and orthogonal Fourier-Mellin moments," *Pattern Recognition*, 35:143–154 (2002).
- [184] Karl, David M. and George A. Knauer. "Vertical distribution, transport, and exchange of carbon in the northeast Pacific Ocean: evidence for multiple zones of biological activity," *Deep-Sea Research*, 31:221–243 (1984).
- [185] Karp-Bass, L., et al. "Optical measurements of phytoplankton size and volume concentration (LISST-100X): Applications and limitations." *Ocean Optics Conference, Montreal, Quebec*. 2006.

- [186] Karp-Bass, Lee, et al. "LISST-100 measurements of phytoplankton size distribution: evaluation of the effects of cell shape," *Limnology and Oceanography: Methods*, 5:396–406 (2007).
- [187] Katkovnik, Vladimir, et al. "Backward discrete wave field propagation modeling as an inverse problem: toward perfect reconstruction of wave field distributions," *Applied Optics*, 48:3407–3423 (2009).
- [188] Khotanzad, Alireza and Yaw Hua Hong. "Invariant Image Recognition by Zernike Moments," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12:489–497 (1990).
- [189] Kienzle, Wolf, et al. "Face Detection - Efficient and Rank Deficient." *Advances in Neural Information Processing Systems 17*. 673–680. MIT Press, 2005.
- [190] Kim, Taegeun and Ting-Chung Poon. "Extraction of 3-D location of matched 3-D object using power fringe-adjusted filtering and Wigner analysis," *Optical Engineering*, 38:2176–2183 (1999).
- [191] Kindermann, Ross and J. Laurie Snell. *Markov Random Fields and Their Applications*. Americal Mathematical Society, 1980.
- [192] Kirby, M. and L. Sirovich. "Application of the Karhunen-Loeve Procedure for Characterization of Human Faces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12:103–108 (1990).
- [193] Kocak, Donna M., et al. "A Focus on Recent Developments and Trends in Underwater Imaging," *Marine Technology Society Journal*, 42-67:52 (2008).
- [194] Köhn, A., et al. "GPU accelerated image registration in two and three dimensions." *British Machine Vision Conference*. 2006.
- [195] Kolmogorov, Vladimir and Ramin Zabih. "What Energy Functions Can Be Minimized via Graph Cuts?," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26:147–159 (2004).

- [196] Kotlyar, V. V., et al. "Simple optical vortices formed by a spiral phase plate," *Journal of Optical Technology*, 74:686–693 (2007).
- [197] Kou, L., et al. "Refractive indices of water and ice in the 0.65 - 2.5 μm spectral range," *Applied Optics*, 32:3531–3540 (1993).
- [198] Kovesi, Peter, "Matlab and Octave Functions for Computer Vision and Image Processing." Centre for Exploration Targeting, School of Earth and Environment, The University of Western Australia. Available at <http://www.csse.uwa.edu.au/~pk/research/matlabfns/>.
- [199] Kovesi, Peter. "Image Features from Phase Congruency," *Videre: Journal of Computer Vision Research*, MIT Press, 1:2–23 (1999).
- [200] Kovesi, Peter. "Edges Are Not Just Steps." *Proceedings of the ACCV2002, The Fifth Asian Conference on Computer Vision, Melbourne, Australia*. 822–827. 2002.
- [201] Kovesi, Peter. "Phase Congruency Detects Corners and Edges." *The Australian Pattern Recognition Society Conference: DICTA, Sydney, Australia*. 309–318. 2003.
- [202] Kreis, Thomas. *Handbook of holographic interferometry: optical and digital methods*. Wiley VCH, 2005.
- [203] Kreis, Thomas M. "Frequency analysis of digital holography," *Opt. Eng.*, 41:771–778 (2002).
- [204] Kreis, Thomas M. "Frequency analysis of digital holography with reconstruction by convolution," *Opt. Eng.*, 41:1829–1839 (2002).
- [205] Kronrod, M. A., et al. "Reconstruction of holograms with a computer," *Soviet Physics: Technical Physics*, 17:333–334 (1972).
- [206] Krotkov, Eric. "Focusing," *International Journal of Computer Vision*, 1:223–237 (1987).

- [207] Kühn, J., et al. “Real-time dual-wavelength digital holographic microscopy with a single hologram,” *Proc. SPIE Optical Measurement Systems for Industrial Inspection*, 6616:661615 (2007).
- [208] Kyrgyzov, I. O., et al. “Kernel MDL to Determine the Number of Clusters.” *Second International Workshop on Machine Learning and Data Mining in Pattern Recognition*. 2001.
- [209] LaBry, Zach A. *Suppression of Thermoacoustic Instabilities in a Swirl Combustor through Microjet Air Injection*. MS thesis, Massachusetts Institute of Technology, 2010.
- [210] Larkin, K. G., et al. “Natural demodulation of two-dimensional fringe patterns. I. General background of the spiral phase quadrature transform,” *Journal of the Optics Society of America A*, 18:1862–1870 (2001).
- [211] Larkin, Kieran G. “Natural demodulation of two-dimensions fringe patterns. II. Stationary phase analysis of the spiral phase quadrature transform,” *Journal of the Optics Society of America A*, 18:1871–1881 (2001).
- [212] Larkin, Kieran G. “Uniform estimation of orientation using local and nonlocal 2-D energy operators,” *Optics Express*, 13:8412 (2005).
- [213] Lee, J. H. and H. T. Kim. “Natural frequency extraction using generalized pencil-of-function method and transient response reconstruction,” *Progress In Electromagnetics Research C*, 4:65–84 (2008).
- [214] Leibe, B., et al. “Coupled Detection and Trajectory Estimation for Multi-Object Tracking.” *Proceedings of the International Conference on Computer Vision, Rio de Janeiro*. 2007.
- [215] Leibe, Bastian, et al. “Segmentation Based Multi-Cue Integration for Object Detection.” *British Machine Vision Conference*. 2006.
- [216] Lenz, P. H., et al. “The need for speed. I. Fast reactions and myelinated axons in copepods,” *Journal of Computational Physiology A*, 186:337–345 (2000).

- [217] Li, Weichang. *Computational Reconstruction of Images from Optical Holograms*. MS thesis, Massachusetts Institute of Technology, 2002.
- [218] Li, Weichang, et al. “Fast Computation of Focal Planes for Sparsely Populated Digital Holograms Using Spectral ℓ_1 Norms.” *Optical Society of America Conference on Digital Holography and Three-Dimensional Imaging, Vancouver, BC*. 2007.
- [219] Li, Weichang, et al. “Focus detection from digital in-line holograms based on spectral ℓ_1 norm and its application to underwater holography.” *IEEE/MTS Oceans Conference Europe, Aberdeen, Scotland*. 2007.
- [220] Li, Weichang, et al. “Focus detection from digital in-line holograms based on spectral ℓ_1 norms,” *Journal of the Optics Society of America A*, 24:3054–3062 (2007).
- [221] Liebling, Michael, et al. “Fresnelets: New Multiresolution Wavelet Bases for Digital Holography,” *IEEE Transactions on Image Processing*, 12:29–43 (2003).
- [222] Liebling, Michael and Michael Unser. “Autofocus for digital Fresnel holograms by use of a Fresnelet-sparsity criterion,” *Journal of the Optics Society of America A*, 21:2424–2430 (2004).
- [223] Lindeberg, Tony. “Feature Detection with Automatic Scale Selection,” *International Journal of Computer Vision*, 30:79–116 (1998).
- [224] Lindeberg, Tony. *Handbook on Computer Vision and Applications*, chapter Principles for Automatic Scale Selection, 239–274. Academic Press, Boston, 1999.
- [225] Liu, Zhiwen, et al. “Holographic recording of fast events on a CCD camera,” *Optics Letters*, 27:22–24 (2002).
- [226] Lohmann, Adolf W., et al. “Space-bandwidth product of optical signals and systems,” *Journal of the Optics Society of America A*, 13:470–473 (1996).
- [227] Loomis, Nick and George Barbastathis. “Space-Bandwidth Transfer in Spherical Reference Holography using Wigner Transforms.” *Digital Holography and Three-Dimensional Imaging, Optical Society of America, Vancouver, Canada*. 2009.

- [228] Loomis, Nick, et al. “A compact, low-power digital holographic imaging system for automated plankton taxonomical classification.” *4th International Zooplankton Production Symposium, Hiroshima, Japan*. 2007.
- [229] Loomis, Nick, et al. “High-speed Phase Recovery Using Chromatic Transport of Intensity Computation in Graphics Processing Units.” *Optics Society of America Digital Holography and 3D Imaging, Miami, FL*. 2010.
- [230] Loomis, Nick C., et al. “Advances in Digital Holography for Real-Time Imaging in Fluid Environments.” *15th International Symposium on Applications of Laser Techniques to Fluid Mechanics, Lisbon, Portugal*. 2010.
- [231] Love, M. S., et al. *The Rockfishes of the Northeast Pacific*. University of California Press, 2002.
- [232] Lowe, David G. “Object recognition from local scale-invariant features,” *Proceedings of the International Conference on Computer Vision*, 2:1150–1157 (1999).
- [233] Lowe, David G. “Distinctive Image Features from Scale-Invariant Keypoints,” *International Journal of Computer Vision*, 60:91–110 (2004).
- [234] Luo, Tong, et al. “Active Learning to Recognize Multiple Types of Plankton,” *Journal of Machine Learning Research*, 6:589–613 (2005).
- [235] Ma, Lihong, et al. “Numerical reconstruction of digital holograms for three-dimensional shape measurement,” *Journal of Optics A: Pure and Applied Optics*, 6:396–400 (2004).
- [236] Mahalanobis, P.C. “On the generalized distance in statistics,” *Proc. National Institute of Science, India*, 2:49–55 (1936).
- [237] Mairesse, Fabrice, et al. “Subpixel determination of imperfect circles characteristics,” *Pattern Recognition*, 41:250–271 (2007).
- [238] Malcolm, James, et al. “A Graph Cut Approach to Image Segmentation in Tensor Space.” *Computer Vision and Pattern Recognition*. 2007.

- [239] Malcolm, James, et al. “Graph cut segmentation with nonlinear shape priors.” *International Conference on Image Processing*. 2007.
- [240] Malkiel, E., et al. “The three-dimensional flow field generated by a feeding calanoid copepod measured using digital holography,” *The Journal of Experimental Biology*, 206:3257–3666 (2003).
- [241] Malkiel, Edwin, et al. “Measurements of plankton distribution in the ocean using submersible holography,” *Measurement Science and Technology*, 10:1142–1152 (1999).
- [242] Martin, Glen and Carol Ness, “Sea of anxiety over rockfish ban: Fishermen, distributors, coastal towns brace for hard times as species recover.” San Francisco Chronicle, June 21 2002.
- [243] Matas, J., et al. “Robust wide-baseline stereo from maximally stable extremal regions,” *Image and Vision Computing*, 22:761–767 (2004).
- [244] Matsushima, Kyoji. “Formulation of the rotational transformation of wave fields and their application to digital holography,” *Applied Optics*, 47:D110–D116 (2008).
- [245] Matsushima, Kyoji, et al. “Fast calculation method for optical diffraction on tilted planes by use of the angular spectrum of plane waves,” *Journal of the Optics Society of America A*, 20:1755–1762 (2003).
- [246] Mattar, M., et al. *Software Tools for Image Analysis*. Technical Report UM-CS-2009-017, Department of Computer Science, University of Massachusetts - Amherst, 2009.
- [247] Matusita, K. “Decision Rules Based on Distance for Problems of Fit, Two Samples and Estimation,” *Ann. Mathematical Statistics*, 26:631–641 (1955).
- [248] McElhinney, Conor P., et al. “Extended focused imaging for digital holograms of macroscopic three-dimensional objects,” *Applied Optics*, 47:D71–D79 (2008).
- [249] McKenna, S. A., et al., “Cruise Report for Finding Coral Expedition.” Living Oceans Society, 2009.

- [250] Mennesson, B., et al. “Use of single-mode waveguides to correct the optical defects of a nulling interferometer,” *Journal of the Optics Society of America A*, 19:596–602 (2002).
- [251] Meyers, Ransom A. and Boris Worm. “Rapid worldwide depletion of predatory fish communities,” *Nature*, 423:280–283 (2003).
- [252] Mézard, Marc. “Where Are the Exemplars?,” *Science*, 315:949–951 (2007).
- [253] Mikkelsen, Ole. personal communications, e-mail, December 10 and 11 2010.
- [254] Milgram, Jerome and Weichang Li. “Computational reconstruction of images from holograms,” *Applied Optics*, 41:853–864 (2002).
- [255] Miller, Daniel J. and Robert N. Lea, “Guide to the Coastal Marine Fishes of California.” State of California Department of Fish and Game, Fish Bulletin 157, 1972.
- [256] Mobley, Curtis D. *Light and Water: Radiative Transfer in Natural Waters*. Academic Press, 1994. Updated version available online and by CD-ROM.
- [257] Moon, Inkyu and Bahram Javidi. “Shape tolerant three-dimensional recognition of biological microorganisms using digital holography,” *Optics Express*, 13:9612–9622 (2005).
- [258] Morgan, Alexia, et al. “Overview of the U.S. east coast bottom longline shark fishery, 1994–2003,” *Marine Fisheries Review*, 71:23–38 (2009).
- [259] Munroe, Randall, “Complex Conjugate.” Available online at <http://xkcd.com/849/>, January 19 2011.
- [260] Musick, J. A., et al. “Marine, Estuarine, and Diadromous Fish Stocks at Risk for Extinction in North America (Exclusive of Pacific Salmonids),” *Fisheries*, 25:6–30 (2000).
- [261] National Commission on the BP Deepwater Horizon Oil Spill and Offshore Drilling, “Deep Water: The Gulf Oil Disaster and the Future of Offshore Drilling,” 2011.

- [262] Nebrensky, J. J., et al. “Grid computing for the numerical reconstruction of digital holograms.” *Proc. SPIE, Photonics Applications in Astronomy, Communications, Industry, and High-Energy Physics Experiments III* 5775. 285–296. 2005.
- [263] Nelleri, Anith, et al. “Recognition and classification of three-dimensional phase objects by digital Fresnel holography,” *Applied Optics*, 45:4046–4053 (2006).
- [264] Nishihara, K., et al. “New method of obtaining particle diameter by the fast Fourier transform pattern of the in-line hologram,” *Optical Engineering*, 36:2429–2439 (1997).
- [265] Office of Naval Research, “Worldwide Ocean Optics Database.” Online at <http://wood.jhuapl.edu/wood/>, accessed March 28 2001.
- [266] Olson, Robert J. and Heidi M. Sosik. “A submersible imaging-in-flow instrument to analyze nano- and microzooplankton: Imaging FlowCytobot,” *Limnology and Oceanography: Methods*, 5:195–203 (2007).
- [267] Onural, L. and P. D. Scott. “Digital decoding of in-line holograms,” *Optical Engineering*, 26:1124–1132 (1987).
- [268] Onural, Levent and Mehmet Tankut Özgen. “Extraction of three-dimensional object-location information directly from in-line holograms using Wigner analysis,” *Journal of the Optics Society of America A*, 9:252–260 (1992).
- [269] Opelt, A., et al. “Weak hypotheses and boosting for generic object detection and recognition.” *Eighth European Conference on Computer Vision, Prague*. 71–84. 2004.
- [270] Oppenheim, Alan A., et al. *Signals and Systems*. Prentice Hall, 1996.
- [271] Orlova, E. L., et al. “The relationship between plankton, copelin, and cod under different temperature conditions,” *ICES Journal of Marine Science*, 62:1281–1292 (2005).
- [272] Owen, Robert B. and Alex A. Zozulya. “In-line digital holographic sensor for monitoring and characterizing marine particulates,” *Optical Engineering*, 39:2187–2197 (2000).

- [273] Pan, Gang and Hui Meng. “Digital holographic PIV for 3D flow measurement.” *ASME International Mechanical Engineering Congress and Exposition, New Orleans, LA*. 2002.
- [274] Pan, Gang and Hui Meng. “Digital holography of particle fields: reconstruction by use of complex amplitude,” *Applied Optics*, 42:827–833 (2003).
- [275] Pantelic, Radosav S., et al. “Bilateral edge filter: Photometrically weighted, discontinuity based edge detection,” *Journal of Structural Biology*, 160:93–102 (2007).
- [276] Pardo, Fernando, et al. “Circle detection and tracking speed-up based on change-driven image processing.” *ICGST International Conference on Graphics, Vision, and Image Processing, Cairo*. 2005.
- [277] Paris, S. and F. Durand. “A Fast Approximation of the Bilateral Filter using a Signal Processing Approach.” *European Conference on Computer Vision, Graz, Austria*. 2006.
- [278] Paris, S. and F. Durand. “A Topological Approach to Hierarchical Segmentation using Mean Shift.” *Proc. IEEE Computer Vision and Pattern Recognition*. 2007.
- [279] Parmesan, Camille and Gary Yohe. “A globally coherent fingerprint of climate change impacts across natural systems,” *Nature*, 421:37–42 (2003).
- [280] Pérez, P., et al. “Color-based probabilistic tracking.” *European Conference on Computer Vision*. 661–675. 2002.
- [281] Pfitsch, D. W., et al. “Development of a free-drifting submersible digital holographic imaging system.” *Proceedings of MTS/IEEE OCEANS 1*. 690–696. 2005.
- [282] Pinnegar, C. R. and L. Mansinha. “Time-local Fourier analysis with a scalable, phase-modulated analyzing function: the S-transform with a complex window,” *Signal Processing*, 84:1167–1176 (2004).
- [283] Pizarro, O., et al. “Relative Pose Estimation for Instrumented, Calibrated Imaging Platforms.” *Proc. VII Digital Computing, Technologies and Applications, Sydney*. 2003.

- [284] Pizarro, O., et al. “Large area 3D reconstructions from underwater surveys.” *Proc. Oceans*. 2004.
- [285] Pizarro, O. and S. B. Williams. “Topic-based habitat classification using visual data.” *Proceedings of IEEE Oceans 2009, Bremen, Germany*. 2009.
- [286] Pomarico, J., et al. “Digital recording and numerical reconstruction of holograms: a new method for displaying light in flight,” *Applied Optics*, 35:8095–8099 (1995).
- [287] Poon, Ting-Chung, editor. *Digital Holography and Three-Dimensional Display: Principles and Applications*. Springer, 2010.
- [288] Pope, R. M. and E. S. Fry. “Absorption spectrum (380-700 nm) of pure water, II. Integrating cavity measurements,” *Applied Optics*, 36:8710–8723 (1997).
- [289] Powlick, J. J., et al. “A retrospective of plankton pumping systems, with notes on the comparative efficiency of towed nets,” *Journal of Plankton Research*, 13:901–912 (1991).
- [290] Pu, Ye and Hui Meng. “Four-dimensional dynamic flow measurement by holographic particle image velocimetry,” *Applied Optics*, 44:7697–7708 (2005).
- [291] Quan, Xiaohong and Edward S. Fry. “Empirical equation for the index of refraction of seawater,” *Applied Optics*, 34:3477–3480 (1995).
- [292] Quinlan, J. R. *Programs for Machine Learning*. Morgan Kaufmann Publishers, 1993.
- [293] Quinn, B. G. and J. M. Fernandes. “A Fast Efficient Technique for the Estimation of Frequency,” *Biometrika*, 78:489–497 (1991).
- [294] Quinn, Barry G. “Recent advances in rapid frequency estimation,” *Digital Signal Processing*, 19:942–948 (2009).
- [295] Rad, Ali Ajdari, et al. “Fast Circle Detection Using Gradient Pair Vectors.” *Proc. VIIth Digital Image Computing, Techniques and Applications, Sydney*. 2003.
- [296] Raskar, Ramesh, et al. “Coded Exposure Photography: Motion Deblurring using Fluttered Shutter.” *Proc. IEEE Computer Vision and Pattern Recognition*. 2006.

- [297] Ray, Sidney F. *Applied photographic optics: lenses and optical systems for photography, film, video, electronic and digital imaging*. Focal Press, 2002.
- [298] Reid, P. C., et al. “The Continuous Plankton Recorder: concepts and history, from Plankton Indicator to undulating recorders,” *Progress in Oceanography*, 58:117–173 (2003).
- [299] Remsen, A., et al. “What you see is not what you catch: A comparison of concurrently collected net, optical plankton counter (OPC), and Shadowed Image Particle Profiling Evaluation Recorder (SIPPER) data from the northeast Gulf of Mexico,” *Deep-Sea Research I*, 51:129–151 (2004).
- [300] Reynolds, R. A., et al. “Measurements and characterization of particle size distributions in coastal waters,” *Journal of Geophysical Research*, 115:C08024 (2010).
- [301] Riebesell, U., et al. “Enhanced biological carbon consumption in a high CO_2 ocean,” *Nature*, 450:545–548 (2007).
- [302] Rife, Jason and Stephen M. Rock. “Segmentation Methods for Visual Tracking of Deep-Ocean Jellyfish Using a Conventional Camera,” *IEEE Journal of Oceanic Engineering*, 28:595–608 (2003).
- [303] Robinson, Gerald H. *Photography, History, and Science*. Carl Mautz Publishing, 2006.
- [304] Rose, George A. *Cod: The Ecological History of the North Atlantic Fisheries*. Breakwater Books, 2007.
- [305] Rothschild, B. J. and T. R. Osborn. “Small-scale turbulence and plankton contact rates,” *Journal of Plankton Research*, 10:465–474 (1988).
- [306] Roy, Shovonlal, et al. “Retrieval of phytoplankton size from bio-optical measurements: theory and applications,” *Journal of the Royal Society Interface* (2010).
- [307] Ruiz, Antonio, et al. “Using Graphics Hardware for Enhancing Edge and Circle Detection.” *Proceedings of the 3rd Iberian Conference on Pattern Recognition and Image Analysis*. 2007.

- [308] Ryle, James P., et al. "Dual wavelength digital holographic Laplacian reconstruction," *Optics Letters*, 35:3018–3020 (2010).
- [309] Rymel, J., et al. "Adaptive eigen-backgrounds for object detection." *International Conference on Image Processing* 3. 24–27. 2004.
- [310] Salah, N., et al. "Application of multiple exposure digital in-line holography to particle tracking in a Bénard-von Kármán vortex flow," *Measurement Science and Technology*, 2008:074001 (19).
- [311] Samson, S., et al. "A System for High-Resolution Zooplankton Imaging," *IEEE Journal of Oceanic Engineering*, 26:671–676 (2001).
- [312] Sandin, S. A., et al. "Baselines and Degradation of Coral Reefs in the Northern Line Islands," *PLoS ONE*, 3:e1548 (2008).
- [313] Schnars, U. and W. Jueptner. *Digital Holography*. Springer, 2005.
- [314] Schnars, U. and W. Jüptner. "Direct recording of holograms by a CCD target and numerical reconstruction," *Applied Optics*, 33:179–181 (1994).
- [315] Schnars, Ulf and W. Jüptner. "Digital recording and reconstruction of holograms in hologram interferometry and shearography," *Applied Optics*, 33:4373–4377 (1994).
- [316] Schnars, Ulf and W. P. O. Jüptner. "Digital recording and numerical reconstruction of holograms," *Measurement Science and Technology*, 13:R85–R101 (2002).
- [317] Schulze, Peter C., et al. "Evaluation of a remotely operated vehicle (ROV) as a tool for studying the distribution and abundance of zooplankton," *Journal of Plankton Research*, 17:1233–1243 (1995).
- [318] Seascan, Inc., "Operation and Maintenance of the Digital Auto VPR," April 18 2008.
- [319] Segelstein, D. J. *The complex refractive index of water*. MS thesis, University of Missouri-Kansas City, 1981.
- [320] Sequoia Scientific, Inc., "Particle Size Analyzers." Online at <http://www.sequoiasci.com/products/Particle.aspx>, December 8 2010.

- [321] Sequoia Scientific, Inc., “LISST-HOLO Submersible Holographic Particle Imaging System.” Online at www.sequoiasci.com/graphics/upload/LHOLO/LISST-HOLO.pdf, accessed March 10 2011.
- [322] Service, Rober F., “Can Sonar Detect Undersea Oil Plumes?.” *Science Insider*, June 22 2010.
- [323] Sharifi, Mohsen, et al. “A Classified and Comparative Study of Edge Detection Algorithms.” *International Conference on Information Technology: Coding and Computing, Las Vegas, NV*. 2002.
- [324] Sheldon, R. W., et al. “The size distribution of particles in the ocean,” *Limnology and Oceanography*, 17:337–340 (1972).
- [325] Sheng, J., et al. “Single beam two-views holographic particle image velocimetry,” *Applied Optics*, 42:235–250 (2003).
- [326] Sheng, J., et al. “Digital holographic microscopy reveals prey-induced changes in swimming behavior of predatory dinoflagellates,” *PNAS*, 104:17512–17517 (2007).
- [327] Sheng, J. and H. Meng. “A genetic algorithm particle pairing technique for 3D velocity field extraction in holographic particle image velocimetry,” *Experiments in Fluids*, 25:461–473 (1998).
- [328] Sheng, Jian, et al. “Digital holographic microscope for measuring three-dimensional particle distributions and motions,” *Applied Optics*, 45:3893–3901 (2006).
- [329] Shepard, Francis. *Submarine Geology*. Harper and Row, 1963.
- [330] Shifrin, K. S. *Physical Optics of Ocean Water*. American Institute of Physics, New York, 1988.
- [331] Shimobaba, T., et al. “Numerical calculation library for diffraction integrals using the graphic processing unit: the GPU-based wave optics library,” *Journal of Optics A: Pure and Applied Optics*, 10:075308 (2008).

- [332] Shimobaba, T., et al. “Real-time digital holographic microscopy using the graphic processing unit,” *Optics Express*, 16:11776–11781 (2008).
- [333] Shnayderman, Aleksander, et al. “A multidimensional image quality measure using singular value decomposition.” *Image Quality and System Performance 5294*, edited by Yoichi Miyake and D. Rene Rasmussen. 82–92. 2003.
- [334] Sieracki, Michael E., et al. “Optical Plankton Imaging and Analysis Systems for Ocean Observation.” *OceanObs, Venice, Italy*. 2009.
- [335] Silfvast, William T. *Laser Fundamentals*. Cambridge University Press, 2000.
- [336] Singh, Dhananjay Kumar and P. K. Panigrahi. “Improved digital holographic reconstruction algorithm for depth error reduction and elimination of out-of-focus particles,” *Optics Express*, 18:2426–2448 (2010).
- [337] Singh, H., et al. “Imaging Coral I: Imaging Coral Habitats with the SeaBED AUV,” *Subsurface Sensing Technologies and Applications*, 5:25–42 (2004).
- [338] Singh, H., et al. “Seabed AUV Offers New Platform for High-Resolution Imaging,” *EOS, Transactions of the AGU*, 85:289, 294–295 (2004).
- [339] Sirovich, L. and M. Kirby. “Low-dimensional procedure for the characterization of human faces,” *Journal of the Optics Society of America A*, 4:519–524 (1987).
- [340] Smith, Julius O. *Spectral Audio Signal Processing, October 2008 Draft*. <http://ccrma.stanford.edu/~jos/sasp/>, accessed December 7, 2010. Online book.
- [341] Smyth, Gordon K. “Employing Symmetry Constraints for Improved Frequency Estimation by Eigenanalysis Methods,” *Technometrics*, 42:277–289 (2000).
- [342] Smyth, Gordon K. “Robust Frequency Estimation Using Elemental Sets,” *Journal of Computational and Graphical Statistics*, 9:196–214 (2000).
- [343] Smyth, W. D. and J. N. Moum. *Encyclopedia of Ocean Sciences*, chapter 3D Turbulence. Academic Press, 2001.

- [344] Soderquist, Peter and Miriam Leeser. “Division and Square Root, Choosing the Right Implementation,” *IEEE Micro*, 17:56–66 (1997).
- [345] Solem, Jan Erik, et al. “Initialization Techniques for Segmentation with the Chan-Vese Model.” *Proceedings of the 18th International Conference on Pattern Recognition* 2. 2006.
- [346] Solomon, S., et al., editors. *Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, The Physical Science Basis*. Cambridge University Press, 2007.
- [347] Solov, A., et al. “Estimating the taxonomic composition of a sample when individuals are classified with error,” *Marine Ecology Progress Series*, 216:309–311 (2001).
- [348] Soria, J. and C. Atkinson. “Towards 3C-3D digital holographic fluid velocity vector field measurement—tomographic digital holographic PIV (Tomo-HPIV),” *Measurement Science and Technology*, 19:074002 (2008).
- [349] Soulez, F., et al. “Inverse problem approach for particle digital holography: accurate location based on local optimization,” *Journal of the Optics Society of America A*, 24:1164–1171 (2007).
- [350] Sourisseau, Marc and Francois Carlotti. “Spatial distribution of zooplankton size spectra on the French continental shelf of the Bay of Biscay during spring 2000 and 2001,” *Journal of Geophysical Research*, 11:C05S09 (2006).
- [351] Sprockett, Daniel D., et al. “Evolutionary analysis of glycosyl hydrolase family 28 (GH28) suggests lineage-specific expansions in necrotrophic fungal pathogens,” *Gene (in press)* (2011).
- [352] Stein, E. and G. Weiss. *Introduction to Fourier Analysis on Euclidean Spaces*. Princeton University Press, NJ, 1971.
- [353] Steinberg, Daniel M., et al. “Towards automous habitat classification using Gaussian Mixture Models.” *IEEE-RSJ International Conference on Intelligent Robots and Systems, Taipei, Taiwan*. 4424–4431. 2010.

- [354] Stern, A. and B. Javidi. “Sampling in the light of Wigner distribution,” *Journal of the Optics Society of America A*, 21:360–366 (2004).
- [355] Stern, A. and B. Javidi. “Sampling in the light of Wigner distribution: errata,” *Journal of the Optics Society of America A*, 21:2038 (2004).
- [356] Stern, A. and B. Javidi. “Improved-resolution digital holography using the generalized sampling theorem for locally band-limited fields,” *Journal of the Optics Society of America A*, 23:1227–1235 (2006).
- [357] Stern, Adrian and Bahram Javidi. “Space-bandwidth conditions for efficient phase-shifting digital holographic microscopy,” *Journal of the Optics Society of America A*, 25:736–741 (2008).
- [358] Stochmal, Wieslaw and Henryk Gurgul. “Optical properties of crude oil - sea water emulsion: comparison of model and experimental results,” *Proc. SPIE Underwater Light Measurements*, 2048:112–122 (1993).
- [359] Stockwell, R. G. “A basis for efficient representation of the S-transform,” *Digital Signal Processing*, 17:371–393 (2007).
- [360] Stockwell, R. G., et al. “Localization of the Complex Spectrum: The S Transform,” *IEEE Trans. Sig. Proc.*, 44:998–1001 (1996).
- [361] Storbeck, Frank and Berent Daan. “Fish species recognition using computer vision and a neural network,” *Fisheries Research*, 51:11–15 (2001).
- [362] Strang, Gilbert and Truong Nguyen. *Wavelets and Filter Banks*. Wellesley-Cambridge Press, 1996.
- [363] Strickler, J. R. and J. S. Hwang. *Focus on Modern Microscopy*, chapter Matched spatial filters in long working distance microscopy of phase objects, 217–239. World Scientific, River Edge, NJ, 2000.
- [364] Strickler, J. R., et al. “Combining Analog Turbulence with Digital Turbulence,” *Scientia Marina*, 61:197–204 (1997).

- [365] Strom-Martin, Virginia, “Economic Decline of North Coast Fisheries: Overview and Strategies for the Future.” California Assembly Select Committee on Rural Economic Development, 1998.
- [366] Sun, H., et al. “Underwater digital holography for studies of marine plankton,” *Philosophical Transactions of the Royal Society A*, 366:1789–1806 (2008).
- [367] Sun, Hongyue, et al. “In Situ Underwater Electronic Holographic Camera for Studies of Plankton,” *IEEE Journal of Oceanic Engineering*, 32:373–382 (2007).
- [368] Szeliski, Richard. *Computer Vision: Algorithms and Applications*. Springer, 2010.
- [369] Tachiki, Mark. L., et al. “Simultaneous depth determination of multiple objects by focus analysis in digital holography,” *Applied Optics*, 47:D144–D153 (2008).
- [370] Tang, Xiaou, et al. “Binary Plankton Image Classification,” *IEEE Journal of Oceanic Engineering*, 31:728–735 (2006).
- [371] Tango, William. “The Circle Polynomials of Zernike and Their Application in Optics,” *Applied Physics A: Materials Science and Processing*, 13:327–332 (1977).
- [372] Teague, Michael R. “Image analysis via the general theory of moments,” *Journal of the Optics Society of America*, 70:920–930 (1980).
- [373] Testorf, Markus and Adolf W. Lohmann. “Holography in phase space,” *Applied Optics*, 47:A70–A77 (2008).
- [374] Tetala, Dattatreya Reddy. *Zooplankton visualization system: design and real-time lossless image compression*. MS thesis, Louisiana State University, 2004.
- [375] Thomas, Ann. *Beauty of Another Order: Photography in Science*. Yale University Press, 1997.
- [376] Thompson, A. B. and R. T. Harrop. “Feeding dynamics of fish larvae on Copepoda in the western Irish Sea, with particular reference to cod *Gadus morhua*,” *Marine Ecology Progress Series*, 68:213–223 (1991).

- [377] Thompson, Brian J., et al. “Application of Hologram Techniques for Particle Size Analysis,” *Applied Optics*, 6:519–526 (1967).
- [378] Thurman, Samuel T. and James R. Fienup. “Phase-error correction in digital holography,” *Journal of the Optics Society of America A*, 25:983–994 (2008).
- [379] Tian, Lei and George Barbastathis. “Digital Holography Applied to Quantitative Measurement of Oil-Drop in Oil-Water Two-Phase Flows.” *Optical Society of America Conference on Digital Holography and Three-Dimensional Imaging, Miami, FL*. 2010.
- [380] Tian, Lei, et al. “Quantitative measurement of size and three-dimensional position of fast-moving bubbles in air-water mixture flows using digital holography,” *Applied Optics*, 49:1549–1554 (2010).
- [381] Tippie, Abbie E. and James R. Fienup. “Phase-error correction for multiple planes using a sharpness metric,” *Optics Letters*, 34:701–703 (2009).
- [382] Tittensor, Derek P., et al. “Global patterns and predictors of marine biodiversity across taxa,” *Nature*, 466:1098–1103 (2010).
- [383] Tkalcic, M. and J. F. Tasic. “Colour spaces: perceptual, historical and applicational background.” *IEEE Region 8 EUROCON 2003, Computer as a Tool, Ljubljana, Slovenia*. 2003.
- [384] Tolimieri, Nick, et al. “Evaluating the SeaBED AUV for Monitoring Groundfish in Untrawable Habitat.” *Marine Habitat Mapping Technology for Alaska*. 2008.
- [385] Torralba, A., et al. “Sharing visual features for multiclass and multiview object detection,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29:854–869 (2007).
- [386] Torralba, Antonio and Aude Oliva. “Statistics of natural image categories,” *Network: Computation in Neural Systems*, 14:391–412 (2003).
- [387] Trefethen, Lloyd N. and David Bau. *Numerical Linear Algebra*. Society for Industrial and Applied Mathematics, Philadelphia, 1997.

- [388] Turk, M. and A. Pentland. “Eigenfaces for Recognition,” *Journal of Cognitive Neuroscience*, 3:71–76 (1991).
- [389] Ujaldón, Manuel, et al. “On the computation of the Circle Hough Transform by a GPU rasterizer,” *Pattern Recognition Letters*, 29:309–318 (2008).
- [390] Ulloa, Osvaldo, et al. “Effect of the particle-size distribution on the backscattering ratio in seawater,” *Applied Optics*, 33:7070–7077 (1994).
- [391] University of Waikato, “Weka 3.” Available online at <http://www.cs.waikato.ac.nz/ml/weka/>.
- [392] Unser, Michael, et al. “Multiresolution Monogenic Signal Analysis Using the Riesz-Laplace Wavelet Transform,” *IEEE Transactions on Signal Processing*, 18:2402–2418 (2009).
- [393] U.S. Commerce Department, “NIST/SEMATECH e-Handbook of Statistical Methods.” available at <http://www.itl.nist.gov/div898/handbook/>, August 02 2008.
- [394] U.S. Department of Commerce, Office of National Marine Sanctuaries, “A Comparison of Seafloor Habitats and Associated Benthic Fauna in Areas Open and Closed to Bottom Trawling Along the Central California Continental Shelf.” Available from the National Marine Sanctuaries Program (www.sanctuaries.nos.noaa.gov), March 2009.
- [395] van de Weijer, J. *Color Features and Local Structure in Images*. PhD dissertation, University of Amsterdam, 2005.
- [396] van de Weijer, J., et al. “Boosting Color Saliency in Image Feature Detection,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28:150–156 (2006).
- [397] van der Schaaf, A. and J. H. van Hateren. “Modelling the Power Spectra of Natural Images: Statistics and Information,” *Vision Research*, 36:2759–2770 (1996).
- [398] Varma, M. and A. Zisserman. “Classifying images of materials: achieving viewpoint and illumination independence.” *European Conference on Computer Vision, Copenhagen*. 255–271. 2002.

- [399] Varma, M. and A. Zisserman. “A statistical approach to material classification using image patch exemplars,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31:2032–2047 (2009).
- [400] Vedaldi, Andrea and Stefano Soatto. “Quick Shift and Kernel Methods for Mode Seeking,” *European Conference on Computer Vision*, 5305:705–718 (2008).
- [401] Vezhnevets, Vladimir and Vadim Konouchine. “‘GrowCut’ - Interactive Multi-Label N-D Image Segmentation By Cellular Automata.” *Proceedings of GraphIcon*. 150–156. 2005.
- [402] Vincent, L. “Morphological grayscale reconstruction in image analysis: applications and efficient algorithms,” *IEEE Transactions on Image Processing*, 2:176–201 (1993).
- [403] Žunić, Joviša, et al. “A Hu moment invariant as a shape circularity measure,” *Pattern Recognition*, 43:47–57 (2010).
- [404] Waller, Laura. *Computational phase imaging based on intensity transport*. PhD dissertation, Massachusetts Institute of Technology, 2010.
- [405] Waller, Laura and George Barbastathis. “Phase from defocused color images.” *Optical Society of America Frontiers in Optics, San Jose, CA*. 2009.
- [406] Waller, Laura, et al. “Phase from chromatic aberration,” *Optics Express*, 18:22817–22825 (2010).
- [407] Wallner, Oswald, et al. “Minimum length of a single-mode fiber spatial filter,” *Journal of the Optics Society of America A*, 19:2445–2448 (2002).
- [408] Walstra, P. “Approximate formulae for the light scattering coefficient of dielectric spheres,” *British Journal of Applied Physics*, 15:1545–1552 (1964).
- [409] Wang, Dayong, et al. “High-fidelity numerical realization of multiple-step Fresnel propagation for the reconstruction of digital holograms,” *Applied Optics*, 47:D12–D20 (2008).

- [410] Wang, Lu, et al. “Background subtraction using incremental subspace learning.” *IEEE International Conference on Image Processing* 5. 45–48. 2007.
- [411] Wang, Zhihui, et al. “Cross-scale, cross-pathway evaluation using an agent-based non-small cell lung cancer model,” *Bioinformatics*, 25:2389–2396 (2009).
- [412] Warner, A. J. and G. C. Hays. “Sampling by the continuous plankton recorder survey,” *Progress in Oceanography*, 1994:237–256 (34).
- [413] Warren, Joseph D. *Estimating Gulf of Maine Zooplankton Distributions Using Multiple Frequency Acoustic, Video and Environmental Data*. PhD dissertation, Massachusetts Institute of Technology and Woods Hole Oceanographic Institution, 2000.
- [414] Watson, J., et al. “The development and sea trials of a subsea holographic camera for large volume in-situ recording of marine organisms.” *Practical Holography XVI and Holographic Materials VIII, Proceedings of the SPIE* 4659. 69–75. 2002.
- [415] Watson, J., et al. “eHoloCam - an electronic holographic camera for subsea analysis.” *IEEE/MTTS OCEANS* 3. 1248–1254. 2004.
- [416] Webb, Geoffrey I. “Decision Tree Grafting From the All-Tests-But-One Partition.” *Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence, Morgan Kaufmann, San Francisco, CA*. 702–707. 1999.
- [417] Wei, Wang, et al. “Image Matching for Geomorphic Measurement Based on SIFT and RANSAC Methods.” *International Conference on Computer Science and Software Engineering*. 2008.
- [418] WET Labs, Inc., “WET Labs In-situ Spectrophotometers.” Online at <http://www.wetlabs.com/products/ac/acall.htm>, accessed December 8 2010.
- [419] White, D. J., et al. “Automated measurement of species and length of fish by computer vision,” *Fisheries Research*, 80:203–210 (2006).
- [420] Wiebe, P. H., et al. “New development in the MOCNESS, an apparatus for sampling zooplankton and micronekton,” *Marine Biology*, 87:313–323 (1985).

- [421] Wiebe, Peter H. and Mark C. Benfield. “From the Hensen net toward four-dimensional biological oceanography,” *Progress in Oceanography*, 56:7–136 (2003).
- [422] Wilder, Kelley. *Photography and Science*. Reaktion Books, 2009.
- [423] Wilkins, M. K. and L. Boddy C. W. Morris. “Proportion estimation with confidence limits,” *Journal of Microbiological Methods*, 43:55–64 (2000).
- [424] Williams, Jerome. *Optical Properties of the Sea*. United States Naval Institute, Annapolis, 1970.
- [425] Wishner, Karen E., et al. “Abundance, distribution and population structure of the copepod *Calanus finmarchicus* in a springtime right whale feeding area in the southwestern Gulf of Maine,” *Continental Shelf Research*, 15:475–507 (1995).
- [426] Woodson, C. B., et al. “Response of copepods to physical gradients associated with structure in the ocean,” *Limnology and Oceanography*, 50:1552–1564 (2005).
- [427] Worm, B., et al. “Rebuilding global fisheries,” *Science*, 325:578–585 (2009).
- [428] Worm, Boris, et al. “Impacts of Biodiversity Loss on Ocean Ecosystem Services,” *Science*, 2006:787–790 (314).
- [429] Worm, Boris and Heike K. Lotze. *Climate Change: Observed Impacts on Planet Earth*, chapter Changes in Marine Biodiversity as an Indicator of Climate Change, 263–279. Elsevier, 2009.
- [430] Worm, Boris and Ransom A. Meyers. “Meta-analysis of cod-shrimp interactions reveals top-down control in oceanic food webs,” *Ecology*, 84:162–173 (2003).
- [431] Wu, Zhongquan, et al. “The Robust Algorithms for Finding the Center of an Arc,” *Computer Vision and Image Understanding*, 62:269–278 (1995).
- [432] Xu, W., et al. “Tracking particles in four dimensions with in-line holographic microscopy,” *Optics Letters*, 28:164–166 (2003).
- [433] Xu, Wenbo, et al. “Digital in-line holography for biological applications,” *Cell Biology*, 98:11301–11305 (2001).

- [434] Yang, M. H., et al. “Detecting Faces in Images: A Survey,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1:34–58 (2002).
- [435] Yang, Weidong, et al. “Phase signature for particle detection with digital in-line holography,” *Optics Letters*, 31:1399–1401 (2006).
- [436] Yi, Wei and S. Marshall. “Circle detection using fast finding and fitting (FFF) algorithm,” *Geo-spatial Information Science*, 3:74–78 (2000).
- [437] Yoder, James A. and Maureen A. Kennelly. “Seasonal and ENSO variability in global ocean phytoplankton chlorophyll derived from 4 years of SeaWiFS measurements,” *Global Biogeochemical Cycles*, 17:1112 (2003).
- [438] Yuen, H. K., et al. “Comparative study of Hough Transform methods for circle finding,” *Image and Vision Computing*, 8:71–77 (1990).
- [439] Zak, Dan, “NOAA ship to search for underwater oil plumes from leak site.” Washington Post, June 2 2010.
- [440] Zhang, J., et al. “Local Features and Kernels for Classification of Texture and Object Categories: A Comprehensive Study,” *International Journal of Computer Vision*, 73:213–238 (2007).
- [441] Zhang, J., et al. “Turbulent flow measurement in a square duct with hybrid holographic PIV,” *Experiments in Fluids*, 23:373–381 (1997).
- [442] Zhao, Feng, et al. “Binary SIPPER plankton image classification using random subspace,” *Neurocomputing*, 73:1853–1860 (2010).
- [443] Zhong, Yu, et al. “Object Tracking Using Deformable Templates,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22:544–549 (2000).
- [444] Zhou, Feng, et al. “Rational phase mask to extend the depth of field in optical-digital hybrid imaging systems,” *Optics Letters*, 34:380–382 (2009).

- [445] Zilniker, Emanuel E. and I. Vaughan L. Clarkson. “Maximum-Likelihood Estimation of Circle Parameters Via Convolution,” *IEEE Transactions on Image Processing*, 15:865–876 (2006).
- [446] Zollett, E. A. “Bycatch of protected species and other species of concern in the US east coast commercial fisheries,” *Endangered Species Research*, 9:49–59 (2009).